PATENT

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

In re Application of:

STEPHEN M. ALLEN ET AL.

CASE NO: BB1157 US CNT

SERIAL NO: 10/659,199

GROUP ART UNIT: 1638

FILED: SEPTEMBER 10, 2003

EXAMINER: KUBELIK, ANNE R.

FOR: A NUCLEIC ACID ENCODING A WHEAT BRITTLE-1 HOMOLOG

## APPEAL BRIEF

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Pursuant to 37 C.F.R. § 1.192, the following is an Appeal Brief in support of the Appeal filed August 24, 2007, appealing the Final Office Action dated February 27, 2007. Submitted herewith is the filing fee for this Appeal Brief in accordance with 37 C.F.R. § 41.20(b)(2). Please charge said fee to Deposit Account No. 501447 (Potter Anderson & Corroon LLP).

## TABLE OF CONTENTS

## I. REAL PARTY IN INTEREST

The real party in interest is E.I. du Pont de Nemours and Company (*hereinafter* "DuPont"), owner of the Application.

## II. RELATED APPEALS AND INTERFERENCES

There are no other appeals or interferences known to Applicants, Applicants' legal representative, or DuPont that will directly affect or be directly affected by or have a bearing on the Board of Patent Appeals and Interferences' (*hereinafter* the "Board") decision in the present Appeal.

## III. STATUS OF THE CLAIMS

Claims 26-29 stand rejected and are the subject of this Appeal. Originally-filed Claims 1-25 have been canceled.

## IV. STATUS OF AMENDMENTS

No amendments were made to the claims in response to the Final Office Action.

## V. SUMMARY OF CLAIMED SUBJECT MATTER

Claim 26, the only independent claim at issue, relates to an isolated polynucleotide comprising: (a) a nucleotide sequence encoding a polypeptide having brittle-1 activity (*see, e.g.*, Applicants' Specification at page 6, lines 17-21), wherein the polypeptide has an amino acid sequence of at least 90% sequence identity when compared to SEQ ID NO:18 (*see, e.g.,* Applicants' Specification at page 8, line 34 – page 9, line 17; page 14, lines 16-36), or (b) a full-length complement of the nucleotide sequence of (a) (*see, e.g.,* Applicants' Specification at page 8, lines 19-33; page 12, lines 3-7).

## VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Whether claims 26-29 are supported by sufficient written description under 35 U.S.C. § 112, 1st Paragraph.

Whether claims 26-29 are enabled under 35 U.S.C. § 112, 1st Paragraph.

## VII. ARGUMENT

### A. Claims 26-29 Comply with the Written Description Requirement of 35 U.S.C. § 112, 1st Paragraph.

Claims 26-29 stand rejected under 35 USC 112, first paragraph, as failing to comply with the written description requirement. During the course of prosecution, the Examiner asserted that "the specification only describes a coding sequence from wheat that comprises SEQ ID NO:17. Applicant does not describe other nucleic acids encompassed by the claims, and the structural and functional features that distinguish all such nucleic acids from other nucleic acids are not provided." June 20, 2005, Non-Final Office Action, at 6 (*hereinafter* "Non-Final OA"). The Examiner further asserted that "[n]o description is provided as to the function of the encoded protein." *Id.*; *see also* March 8, 2006, Final Office Action, at 8 (*hereinafter* "First Final OA"). The Examiner thus concluded that "Applicant has not, in fact, described nucleic acids that encode a protein with 90% identity to SEQ ID NO:18 within the full scope of the claims, and the specification fails to provide an adequate written description of the claimed invention." Non-Final OA, at 6; *see also* First Final OA, at 9.

Applicants' claimed invention, however, substantially conforms to Example 14 of the "Synopsis of Application of Written Description Guidelines", 66 Fed. Reg. 1099 (Jan. 5, 2001), *available at* http://www.uspto.gov/web/menu/written.pdf (last visited Oct. 5, 2007) (*hereinafter* "Written Description Guidelines"). In Example 14, the exemplary claim is directed to "A protein having SEQ ID NO:3 and variants thereof that are at least 95% identical to SEQ ID NO:3 and catalyze the reaction of A→B." Written Description Guidelines, at 53. Included in the Example 14 specification is an "indicat[ions] that procedures for making proteins with substitutions, deletions, insertions and additions is routine in the art and . . . an assay for detecting the catalytic activity of the protein." *Id.* Under the "Analysis" section of Example 14, the requirements of 95% identity to SEQ ID NO:3 and having catalytic activity "are essential to the operation of the claimed invention." *Id.* The procedures of making and testing sequences having 95% identity to SEQ ID NO:3 are determined to be "conventional." *Id.* Example 14 concludes that

> [t]he single species disclosed is representative of the genus because all members have at least 95% structural identity with the reference compound and because of the presence of an assay which applicant provided for identifying all of the at least 95% identical variants of SEQ ID NO:3 which are capable of the specified catalytic activity. One of skill in the art would conclude that applicant was in possession of the necessary common attributes possessed by the members of the genus.

*Id.* at 54-55.

Applicants' claimed invention, though directed to the nucleotide sequences encoding the proteins having brittle-1 activity, is structured similarly to that of the Example 14 claim. The claimed nucleotide sequences encode proteins having 90% identity to SEQ ID NO:18, with the encoded proteins having brittle-1 activity. Like Example 14, there is not substantial variation in the encoded proteins, because the entire genus must have 90% sequence identity to SEQ ID NO:18 and have brittle-1 activity. Procedures for producing proteins having 90% identity to SEQ ID NO:18 are well-known in the art. For example, alterations in a nucleic acid fragment that result in the production of a chemically equivalent amino acid at a given site, but do not affect the functional properties of the encoded polypeptide, are well known in the art, and such alterations can be obtained by, for example, site directed mutagenesis as described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (cited in Applicants' Specification at page 14, lines 9-12). Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes that result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the polypeptide molecule would also not be expected to alter the activity of the polypeptide. *See, e.g.,* Applicants' Specification, page 7, line 29 – page 8, line 9. Alterations in nucleic acid fragments can also be introduced by error-prone polymerase chain reaction (PCR), a technique that is well known by those skilled in the art. *See, e.g.,* U.S.

Patent Nos. 4,683,195 and 4,800,159; *see also* You L. & Arnold F.H., Protein Eng. 9:77-83 (1996). Another technique for obtaining homologous proteins is directed molecular evolution. Unlike protein engineering, in which proteins are improved by making specific changes to them, directed evolution involved mutating genes *in vitro* and screening the resulting proteins for desired activity. *See* Crameri A. *et al.*, Nat. Biotech. 15:436-38 (1997). Other techniques are well known in the art, but Applicants refrain from providing such information as it would be merely cumulative of the information provided above.

Further, Applicants provided a brittle-1 assay from Shannon *et al.*, Plant Physiol. 117:1235-52 (1998), used to identify the proteins having 90% sequence identity to SEQ ID NO:18 that also have brittle-1 activity. The assay in Shannon *et al.* consists of measuring [$^{14}$C]Glucose uptake from ADP-Glucose (ADP-Glc) and incorporation into methanol- and water-insoluble products as described on page 1239 and Table V of Shannon *et al.* In short, amyloplasts isolated from endosperm tissue are added to a reaction mixture (200 µl, final volume) containing 100 mM Bicine, pH 8.5, 0.5 M sorbitol, 12.5 mM EDTA, 10 mM GSH, 50 mM potassium acetate, and 4 mM [$^{14}$C]ADP-Glc plus 60 or 80 µl of amyloplast fraction. Amyloplasts are isolated from endosperm tissue of kernels aged between 10 to 14 days past pollination.

It is well-established that an applicant need not disclose that which is known in the art. *In re Buchner*, 929 F.2d 660, 661, 18 USPQ2d 1331, 1332 (Fed. Cir. 1991). Applicants' citation of this reference in the specification should thus be sufficient evidence of written description of the assay.

Applicants thus respectfully submit, in accordance with Example 14 of the Written Description Guidelines, that the claimed invention is supported by sufficient written description in Applicants' Specification.

The Board's recent decision in *Ex parte Kubin*, Appeal 2007-0819 (BPAI May 31, 2007), *available at* http://www.uspto.gov/web/offices/dcom/bpai/prec/fd070819.pdf (last visited Oct. 9, 2007), is distinguishable. In *Kubin*, the claim at issue was directed to "An isolated nucleic acid molecule comprising a polynucleotide encoding a polypeptide at least 80% identical to amino acids 22-221 of SEQ ID NO:2, wherein the polypeptide binds CD48." *Kubin*, Appeal 2007-0819, at 3. The

*Kubin* specification "does not disclose any variants in which the nucleotide sequence encoding amino acids 22-221 of SEQ ID NO:2 is varied." *Id.* at 13 (emphasis added). Further, the Board noted that there was no disclosure of "correlation between function (binding to CD48) and structure responsible for binding to CD48 (other than the entire extracellular domain) such that the skilled artisan would have known what modifications could be made . . . without losing function." *Id.* In light of these facts, the Board concluded that "[p]ossession may not be shown by merely describing how to obtain possession of members of the claimed genus or how to identify their common structural features." *Id.* at 16 (citing *Univ. of Rochester v. GD Searle & Co.*, 358 F.3d 916, 927, 69 USPQ2d 1886, 1895 (Fed. Cir. 2004)). Of particular concern was Appellants failure to "describe[ ] what domains of [SEQ ID NO:2] are correlated with the required binding to CD48, and thus [Appellants] have not described which . . . amino acids can be varied and still maintain binding." *Id.*

Applicants' Specification, however, provides sufficient guidance as to what amino acids could be modified without affecting brittle-1 activity. While the Examiner has focused on lack of disclosure of sequences having 90% plus identity to SEQ ID NO:18, Figure 1 of Applicants' Specification provides a sequence comparison between SEQ ID NO:18 and a known brittle-1 protein (SEQ ID NO:21) that has only 57.3% identity to the claimed sequence, which provides a clear picture of regions of brittle-1 proteins that have high homology, and are thus likely more susceptible to modification, and regions having low or no homology where more modifications can be made. For example, at the N-termini of these sequences, ten of the first 11 amino acids are identical, with the lone difference being a conservative amino acid substitution of the valine at amino acid seven of SEQ ID NO:21 to an alanine in SEQ ID NO:18. *See, e.g.*, Wu T.D. & Brutlag D.L., Proc. Int. Conf. Intell. Syst. Mol. Biol. 4:230-40 (1996). To the skilled artisan, the significant sequence identity at the N-terminus indicates that little or no sequence substitution should be made there, and if made that conservative substitutions would be preferred, in order to maintain brittle-1 activity. Another example of high homology is amino acids 137-219 of SEQ ID NO:21 and amino acids 125-207 of SEQ ID NO:18. Of these 83 amino acids, only six are different. Four of these substitutions are conservative (two glutamines to arginines, an asparagine to threonine, and a phenylalanine to tyrosine), while two

are non-conservative (an isoleucine to serine and a threonine to proline). Other regions of high homology, for example amino acids 228-418 of SEQ ID NO:21 and amino acids 216-404 of SEQ ID NO:18, provide further guidance as to where and what type of substitution could be made.

By contrast, the C-termini of these proteins are significantly different. After amino acid 404 of SEQ ID NO:18 and amino acid 418 of SEQ ID NO:21, not only does SEQ ID NO:18 contain 11 additional amino acids compared to SEQ ID NO:21, but there is also very little sequence homology between the two sequences. Thus, the skilled artisan could expect that amino acid substitutions, deletions, and/or additions in this region would have little effect on brittle-1 activity as compared to, for example, the N-terminal regional. Another region of low sequence homology can be found at amino acids 54-136 of SEQ ID NO:21 and amino acids 61-124 of SEQ ID NO:18. Similar to the C-termini of SEQ ID NOs: 18 and 21, there is a significant difference in amino acid count in this region (83 amino acids for SEQ ID NO:21 and 64 amino acids for SEQ ID NO:18). The skilled artisan could thus conclude that this region of brittle-1 proteins can have significant amino acid substitutions, deletions, and/or substitutions yet still retain brittle-1 activity. Applicants thus submit that the specification provides sufficient guidance as to what regions of SEQ ID NO:18 could be modified and in what way to produce a protein having (1) at least 90% identity to SEQ ID NO:18 and (2) have brittle-1 activity.

Consequently, *Kubin* does not compel a result of lack of written description here and, indeed, should support Applicants' assertion of adequate written description for the claimed invention because Applicants have described a correlation between structure and function for SEQ ID NO:18. *Cf. Kubin*, Appeal 2007-0819, at 17 ("Without a correlation between structure and function, [Appellants'] claim does little more than define the claimed invention by function.").

Similarly, the cases cited by the Examiner in the Non-Final OA in support of the written description rejection do not compel a result of insufficient written description. The Examiner cites *Regents of the University of California v. Eli Lilly & Co.*, 119 F.3d 1559, 43 USPQ2d 1398 (Fed. Cir. 1997), for the principle that "[t]he name cDNA is not itself a written description of that DNA; it conveys no distinguishing information concerning its identity." Non-Final OA, at 6 (quoting *Eli*

*Lilly*, 119 F.3d at 1567, 43 USPQ2d at 1405). In *Eli Lilly*, the patentee described one mammalian insulin cDNA (and thus one vertebrate insulin cDNA), rat insulin cDNA. *Eli Lilly*, 119 F.2d at 1568, 43 USPQ2d at 1405. One generic claim at issue in *Eli Lilly* covered all mammalian insulin cDNAs, with no limitation as to identity to the disclosed rat sequence. *Id.* at 1563, 1401. The Federal Circuit thus concluded, as quoted by the Examiner, that

> a generic statement such as 'vertebrate insulin cDNA' or 'mammalian insulin cDNA,' without more, is not an adequate written description of the genus because it does not distinguish the claimed genus from others, except by function. It does not specifically define any of the genes that fall within its definition. It does not define any structural features commonly possessed by members of the genus that distinguish them from others. One skilled in the art therefore cannot as one can do with a fully described genus, visualize or recognize the identity of the members of the genus. A definition by function, as we have previously indicated, does not suffice to define the genus because it is only an indication of what the gene does, rather than what it is.

Non-Final OA, at 6 (quoting *Eli Lilly*, 119 F.3d at 1568, 43 USPQ2d at 1406). Here, Applicants claimed invention is distinguished from other "genera" of brittle-1 proteins by claimed identity to SEQ ID NO:18. All genes covered by the claimed genus are instantly recognizable from those not covered by the claim. The claimed genus is not defined solely by function but, as described above, is defined by the structure of SEQ ID NO:18 and its relationship to known brittle-1 proteins. The Examiner's reliance on *Eli Lilly* is thus misplaced.

The other case that the Examiner cited in the Non-Final OA was *Amgen, Inc. v. Chugai Pharm. Co.*, 927 F.2d 1200,18 USPQ2d 1016 (Fed. Cir. 1991), for the principle that "[i]t is not sufficient to define [a gene] solely by its principal biological property, *e.g.*, encoding human erythropoietin, because an alleged conception having no more specificity than that is simply a wish to know the identity of any material with that biological property." Non-Final OA, at 7 (quoting *Amgen*, 927 F.2d at 1206, 18 USPQ2d at 1021). While Applicants do not question the propriety of this statement, it relates to conception, not written description. Applicants believe that the skilled artisan would not question that Applicants conceived the claim 26 invention. In the statement following the Examiner's quote from *Amgen*, the Federal Circuit held that "when an inventor is unable to envision the detailed constitution of a

gene so as to distinguish it from other materials, as well as a method for obtaining it, conception has not been achieved until reduction to practice has occurred, *i.e.*, until after the gene has been isolated." *Amgen*, 927 F.2d at 1206, 18 USPQ2d at 1021. Under this test, Applicants have (1) detailed the constitution of SEQ ID NO:17 and its variants that encode polypeptides having at least 90% identity to SEQ ID NO:18 so as to distinguish these brittle-1 genes from the prior art, (2) have detailed numerous methods for obtaining the claimed genes (*see, e.g.*, page 15, line 11 – page 16, line 22), and (3) in any event have isolated SEQ ID NO:17 from maize (that is, reduced it to practice). Thus, Applicants conceived the claimed invention.

Other recent Federal Circuit cases support Applicants' position that the claims are supported by sufficient written description. In *Invitrogen Corp. v. Clontech Labs., Inc.*, 429 F.3d 1052, 77 USPQ2d 1161 (Fed. Cir. 2005), the claimed invention related to modified reverse transcriptases (RTs) having DNA polymerase activity and reduced RNase H activity, wherein the RTs were encoded by nucleotide sequences derived from a retrovirus, yeast, Neurospora, Drosophila, primates, or rodents. *Invitrogen*, 429 F.3d at 1072, 77 USPQ2d at 1174. The unmodified RTs from this list of organisms were known, and the specification provided test data for the one disclosed modified RT (that is, the only disclosed sequence actually covered by the claim). *Id.* In affirming the district court's finding of sufficient written description, the Federal Circuit distinguished *Eli Lilly* and *Fiers v. Revel*, 984 F.2d 1164, 25 USPQ2d 1601 (Fed. Cir. 1993), noting that "[i]In those cases, the patent specifications at issue did not identify the sequence (structure) of any embodiment of DNA claimed therein." *Invitrogen*, 429 F.3d at 1073, 77 USPQ2d at 1175-76. Here, Applicants claimed genus is of a similar, and probably smaller, scope as that in *Invitrogen*.[1] Applicants' Specification discloses SEQ ID NO:18 (equivalent to the known RT sequences), and Applicants claim a genus having identity thereto in combination with brittle-1 function. While Applicants did not provide test data, the lack of working examples should not affect the written description analysis. *See Falkner v. Inglis*, 448 F.3d 1357, 1366, 79 USPQ2d 1001, 1007 (Fed. Cir. 2006) (confirming that "examples are not

[1] The scope of the *Invitrogen* genus covers any modification to any RT from any of the claimed organisms, so long as the RT has DNA polymerase activity and reduced RNase H activity.

necessary to support the adequacy of a written description . . ." (emphasis in original)).

In *In re Wallach*, 378 F.3d 1330, 71 USPQ2d 1939 (Fed. Cir. 2004), the applicants there disclosed actual possession of a tumor necrosis factor binding protein II ("TBP-II"). *Wallach*, 378 F.3d at 1331, 71 USPQ2d at 1940. From the isolated TBP-II protein, the applicants obtained the N-terminal portion (the first ten amino acids) of the isolated protein. *Id.* The applicants demonstrated that the isolated TBP-II inhibited the cytotoxic effect of tumor necrosis factor. *Id.* From this disclosure, the applicants claimed isolated polynucleotides encoding TBP-II. *Id.* at 1331-32, 1940. In finding lack of written description for the polynucleotide claims, the Federal Circuit first noted that

> the complete amino acid sequence of a protein may put one in possession of the genus of DNA sequences encoding it, and that one of ordinary skill in the art . . . may have therefore been in possession of the entire genus of DNA sequences that can encode the disclosed partial protein sequence, even if individual species within that genus might not have been described or rendered obvious.

*Id.* at 1333, 1942. The *Wallach* applicants failed this test because, while they may have had actual possession of TBP-II, "possession of the protein says nothing about whether they were in possession of the protein's amino acid sequence." *Id.* at 1334, 1943. Applicants here, however, demonstrated possession of the amino acid sequences covered encoded by the claimed polynucleotides, and, as noted in *Wallach*, Applicants should not be required to "list every possible permutation of the nucleic acid sequences that can encode a particular protein for which the amino acid sequence is disclosed. . . ." *Id.* at 1334, 1942.

As noted in *Capon v. Eshhar*, 418 F.3d 1349, 76 USPQ2d 1078 (Fed. Cir. 2005) "[p]recedent distinguishes among generic inventions that are adequately supported, those that are merely a 'wish' or 'plan,' the words of *Fiers v. Revel*, 984 F.2d at 1171, and those in between, as illustrated by *Noelle v. Lederman*, 355 F.3d [1343, 1350, 69 USPQ2d 1508 (Fed. Cir. 2004)]; the facts of the specific case must be evaluated." *Capon*, 418 F.3d at 1360, 76 USPQ2d at 1086. In *Fiers*, one party in an interference claimed that disclosure of how a skilled artisan could, by reverse transcription, obtain an undisclosed DNA sequence was sufficient to support written

description of the count.[2] *Fiers*, 984 F.2d at 1170, 25 USPQ2d at 1605-06. The Board disagreed, and the Federal Circuit affirmed, holding that "[a]n adequate written description of a DNA requires more than a mere statement that it is part of the invention and reference to a potential method for isolating it; what is required is a description of the DNA itself." *Id.* The specification at issue "just represent[ed] a wish, or arguably a plan, for obtaining the DNA." *Id.* at 1171, 1606. Conversely, Applicants' claimed invention cannot be considered a mere wish or plan, because Applicants disclosed the sequence of wheat brittle-1 protein, SEQ ID NO:18, along with the actual nucleotide sequence that encodes the protein, SEQ ID NO:17. Applicants thus provided a description of the DNA itself. In *Noelle*, the patentee claimed human CD40CR antibodies on the basis of a American Type Culture Collection deposit of a hybridoma that secreted the mouse form of the antibody. *Noelle*, 355 F.3d at 1349, 69 USPQ2d at 1514. The patentee failed to disclose any structural elements of the mouse antibody and failed to disclose a fully characterized antigen. *Id.* The Federal Circuit found this lack of disclosure fatal to the patentee's asserted written description sufficiency, holding that no structural elements for the antibody were disclosed in *Noelle* specification. *Id.* at 1349-50, 1514 (noting that the patentee "attempted to define an unknown by its binding affinity to another unknown"). Though not explicitly stated in *Noelle*, the court implied that deposit of the hybridoma was sufficient disclosure of the mouse antibody in accordance with *Enzo Biochem, Inc. v. Gen-Probe Inc.*, 323 F.3d 956, 964, 63 USPQ2d 1609, 1613 (Fed. Cir. 2002). *Noelle*, 355 F.3d at 1349, 69 USPQ2d at 1514. Unlike *Noelle*'s "in between" position on the written description spectrum, Applicants' Specification includes description of brittle-1 structural elements, that is, the actual wheat sequence set forth in SED ID NO:18. Because Applicants' Specification contains structural disclosure beyond that of *Fiers* (only a wish or plan there) and *Noelle* (description of mouse antibody only through deposit of hybridoma producing that antibody), Applicants respectfully submit that their Specification describes the claimed genus of brittle-1-encoding polynucleotides.

[2] The count read: "A DNA which consists essentially of a DNA which codes for a human fibroblast interferon-beta polypeptide." *Fiers*, 984 F.2d at 1166, 25 USPQ2d at 1603.

In light of the above arguments, Applicants respectfully request withdrawal of the rejections of claims 26 and 30-40 under 35 U.S.C. § 112, first paragraph, written description.

**B. Claims 26-29 are Enabled Under 35 U.S.C. § 112, 1st Paragraph.**

Claims 26-29 stand rejected under 35 USC 112, first paragraph, because the Applicants' Specification while being enabling for nucleic acids encoding a SEQ ID NO:18 and constructs and vectors comprising them, allegedly does not reasonably provide enablement for nucleic acids encoding a protein with 90% identity to SEQ ID NO:18 and constructs and vectors comprising them. During prosecution, the Examiner asserted that "[t]he instant specification fails to provide guidance for how to make or isolate nucleic acids encoding proteins with 90% identity to SEQ ID NO:18—specific hybridization or PCR conditions, probes or primers are not recited." Non-Final OA, at 3. Further, the Examiner asserted that "[t]he instant specification fails to teach essential regions of the encoded protein." *Id.* at 3-4. The Examiner also stated that

> [t]he instant specification fails to provide guidance for which amino acids of SEQ ID NO:18 can be altered and to which other amino acids, and which amino acids must not be changed, to maintain adenylate translocator activity of the encoded protein. The specification also fails to provide guidance for which amino acids can be deleted and which regions of the protein can tolerate insertions and still produce a functional enzyme.

*Id.* at 4. The Examiner thus concluded that, because Applicants' Specification allegedly does not describe transformation of plants with a gene encoding a protein having 90% identity to SEQ ID NO:18, "undue trial and error experimentation would be required to screen through the myriad of nucleic acids encompassed by the claims and plants transformed therewith, to identify those with altered starch, if such plants are even obtainable." *Id.* at 5.

Applicants agree with the Examiner that a specification must enable one of ordinary skill in the art to make and use the claimed invention without undue experimentation. Applicants respectfully submit, however, that the Examiner's conclusion of nonenablement of sequences having 90% identity to SEQ ID NO:18 is erroneous because any experimentation needed to practice the present invention

would be routine. "[A] patent specification complies with the statute even if a 'reasonable' amount of routine experimentation is required in order to practice a claimed invention, but that such experimentation must not be 'undue.'" *Enzo Biochem, Inc. v. Calgene, Inc.*, 188 F.3d 1362, 1371, 52 USPQ2d 1129, 1135 (Fed. Cir. 1999). Factors to consider when deciding whether experimentation is undue include: "(1) the quantity of experimentation necessary, (2) the amount of direction or guidance presented, (3) the presence or absence of working examples, (4) the nature of the invention, (5) the state of the prior art, (6) the relative skill of those in the art, (7) the predictability or unpredictability of the art, and (8) the breadth of the claims." *In re Wands*, 858 F.2d 731, 737, 8 USPQ2d 1400, 1404 (Fed. Cir. 1988). Applicants address each of the *Wands* factors below.

(1) The quantity of experimentation needed is quite low. As noted above, methods of producing nucleotide sequences are well-known in the art. A method of testing for brittle-1 activity cited in Applicants' Specification is known in the art and is summarized in the Written Description section of this brief.

(2) Applicants' Specification provides sufficient direction for producing nucleotide sequences encoding proteins having 90% identity to SEQ ID NO:18 and a specific assay for brittle-1 activity is provided.

(3) Applicants admit that there are no working examples showing brittle-1 activity in Applicants' Specification. A specification's lack of working examples, however, does not automatically equate to nonenablement of the claimed invention.

> A claim will not be invalidated on section 112 grounds simply because the embodiments of the specification do not contain examples explicitly covering the full scope of the claim language. That is because the patent specification is written for a person of skill in the art, and such a person comes to the patent with the knowledge of what has come before.

*LizardTech, Inc. v. Earth Resource Mapping, Inc.*, 424 F.3d 1336, 1345, 76 USPQ2d 1724, 1733 (Fed. Cir. 2005) (internal citation omitted).

(4) The invention is one of nucleotide sequences encoding proteins having brittle-1 activity. Such an invention requires some experimentation for even routine techniques.

(5) Brittle 1 protein (BT1), the major protein in amyloplast envelope membranes of wild-type maize endosperm, is required for normal levels of starch

accumulation in maize kernels. Loss of BT1 in the brittle1 (bt1) mutant results in an 80% reduction in kernel starch. The relation of BT1 to starch accumulation was investigated in two well-characterized maize endosperm suspension-cultured cell lines which were derived from 10 days post pollination (DPP) kernels of inbred A636 and 12-DPP kernels of the waxy mutant in the A636 background. Starch in A636 endosperm cultures accounted for about 1.5% of the fresh weight of cells but BT1 was not detectable in amyloplast membranes or in microsomal membranes isolated from the cultured cells. *See* Cao H. & Shannon, J.C., Physiol. Plantarum 97:665-73 (1996).

*In vivo* studies of carbohydrate metabolism during development of maize kernels showed that bt1 kernels accumulated more than 13 times as much adenosine 5'-diphosphoglucose (ADP-Glc) as normal kernels. Activity of starch synthase in bt1 endosperm was equal to that in endosperm extracts from normal kernels. The ADP-Glc accumulation in bt1 endosperm cells was not due to a deficiency in starch synthase, and results indicated that AGPase is the predominant enzyme responsible for *in vivo* synthesis of ADP-Glc in bt1 mutant kernels. *See* Shannon J.C. *et al.*, Plant Physiol. 110:835-43 (1996).

(6) This invention is related to the biotechnical arts in a well-known pathway, the transport of ADP-glucose from the cytosol to the plastid, and the skill level of the artisan is very high. *See, e.g., Kubin*, Appeal 2007-0819, at 14 (noting that the level of skill in the molecular biology art is high). The skilled artisan is therefore very familiar with the pathway and well versed in many methods and techniques of, for example, gene manipulation, protein synthesis, and enzyme action.

(7) Claim 26 is directed to a nucleotide sequence encoding a protein having a specified activity. It is unreasonable for Applicants to provide a cookbook recipe of how to practice the invention. Rather, Applicants have depended on the skill and experience of the skilled artisan to implement the invention using nucleotide sequences encoding polypeptides having brittle-1 activity. Applicants expect that the skilled artisan would be aware of successful molecular biology and biochemistry methods and therefore be capable of producing the described sequences and testing these sequences for brittle-1 activity.

(8) The Examiner's concerns about the number of possible sequences having 90% identity to SEQ ID NO:18 are unfounded. *See, e.g.,* Second Final OA, at 4 ("Making all possible single amino acid substitutions in an [sic] 432 amino acid long protein . . . would require making and analyzing $19^{432}$ ($2.6 \times 10^{552}$) nucleic acids; these proteins would have 99.8% identity to SEQ ID NO:18."). Indeed, the number of possible claimed sequences should not itself form the basis of an enablement rejection. *See, e.g., Novozymes A/S v. Genencor Int'l, Inc.*, 446 F. Supp. 2d 297, 330 (D. Del. 2006) (noting that, with claims 95% identity to a disclosed sequence, a "large number [of possible sequences] alone is not sufficient to show a lack of enablement . . .").

Outside of factor (3), the *Wands* factors support Applicants assertion that any experimentation required to practice the present claims would be routine. "It is well established that a patent applicant is entitled to claim his invention generically when he describes it sufficiently to meet the [enablement requirement]." *Amgen*, 927 F.2d at 1213, 18 USPQ2d at 1027. In *Amgen*, the court found that a generic claim covering all possible DNA sequences encoding any polypeptide having an amino acid sequence "sufficiently duplicative" of erythropoietin ("EPO") and which causes bone marrow cells to increase production of reticulocytes and red blood cells, and increases hemoglobin synthesis or iron uptake as being nonenabled where the patentee only provided information "of how to make EPO and very few analogs." *Id.* at 1213-14, 1027. As noted in *Novozymes*, however, "[t]he problem in [*Amgen*] was that the claim scope covered <u>any</u> gene that could be used to express proteins of various sizes that had one or more of the biological properties of EPO." *Novozymes*, 446 F. Supp. 2d at 330 (emphasis added). Unlike the patentee in *Amgen*, Applicants are not claiming <u>all</u> nucleotide sequences encoding enzymes having brittle-1 activity but merely those having 90% identity to SEQ ID NO:18. Even the *Amgen* court recognized that the enablement requirement should not be extended beyond reasonableness when it noted that the disclosure at issue there might have been sufficient to enable a claim for EPO analogs similar to those described in that specification. *Amgen*, 927 F.2d at 1213, 18 USPQ2d at 1027 (noting that the patentee's "disclosure might well justify a generic claim encompassing these and

similar analogs, but it represents inadequate support for [patentee's] desire to claim all EPO gene analogs").

Applicants' situation is similar to that in *Novozymes*. There, patentee's claim 1 read:

> A variant of a parent *Bacillus stearothermophilus* alpha-amylase, wherein the variant has an amino acid sequence which has at least 95% homology to the parent *Bacillus stearothermophilus* alpha-amylase and comprises a deletion of amino acids 179 an [sic] 180, using SEQ ID NO:3 for numbering, and wherein the variant has alpha-amylase activity.

*Novozymes*, 446 F. Supp. 2d at 306. *Novozymes* concluded that "requiring at least 95% homology with [the identified sequence] makes the variants sufficiently similar so that the enablement requirement is satisfied. By contrast to *Amgen*, the claim scope [in *Novozymes*] is limited quantitatively to similarity between protein sequences and not just to a requirement for alpha-amylase-like activity." *Novozymes*, 446 F. Supp. 2d at 300. Applicants' current claims are similarly structured. There is a quantitative limit to the similarity between SEQ ID NO:18 and other proteins in the claimed genus and all proteins having 90% identity to SEQ ID NO:18 must have brittle-1 activity. Thus, Applicants' claimed invention should be sufficiently enabled.

Further evidencing enablement of claim 26 is that the novel aspect of the invention is enabled in Applicants' Specification. In a recent Federal Circuit case, the court clarified that "[a]lthough the knowledge of one skilled in the art is indeed relevant [to an enablement determination], the novel aspect of an invention must be enabled in the patent." *Auto. Techs. Int'l, Inc. v. BMW of N. Am., Inc.*, 2007 U.S. App. LEXIS 21271, at *22 (Fed. Cir. Sept. 6, 2007). In the present application, the novel aspect of the invention is the sequence set forth in SEQ ID NO:18 and variants thereof. As SEQ ID NO:18 was present in the sequence listing, which is considered part of the specification as filed, the novel aspect of the invention is enabled in the specification. Whether or not the claimed sequences have brittle-1 activity is irrelevant to the novelty of the claimed sequences; a claim directed solely to "a nucleotide sequence encoding an amino acid sequence having 90% identity to SEQ ID NO:18" would be novel without the brittle-1 activity limitation, which is present for

section 112 purposes only. Indeed, brittle-1 activity itself is not novel; as Applicants' Specification notes,

> [b]rittle-1 is one of several corn genes that, when mutated, cause the accumulation of sugars, rather than starch, in developing corn seeds. It has been shown that the brittle-1 gene encodes a plastidic membrane transporter that is involved in the transport of ADP-glucose from the cytosol to the plastid where it is used for starch biosynthesis.

Applicants' Specification, at page 1, lines 12-15 (citation omitted). Therefore, this knowledge can be imputed from those skilled in the art to supplement the present disclosure, as routine experimentation (a brittle-1 assay) provides the determination of whether a sequence having 90% identity to SEQ ID NO:18 is within the scope of the claim 26 invention. *See Invitrogen*, 429 F.3d at 1070-71, 77 USPQ2d at 1173 ("The scope of enablement . . . is that which is disclosed in the specification plus the scope of what would be known to one of ordinary skill in the art without undue experimentation." (quoting *Nat'l Recovery Techs., Inc. v. Magnetic Separation Sys., Inc.*, 166 F.3d 1190, 1194, 49 USPQ2d 1671, 1674 (Fed. Cir. 1999))).

*Kubin* also supports Applicants' enablement arguments. As noted above, the *Kubin* Appellants claimed polynucleotides encoding polypeptides having 80% identity to a defined amino acid sequence, which has a defined binding activity. *Kubin*, Appeal 2007-0819, at 3. In Appellants' specification, the Board found that Appellants' specification taught how to make variants of the defined amino acid sequence, how to calculate identity between the defined amino acid sequence and the variants, and how to test the variant for the claimed binding activity. *Id.* at 13. The specification did not disclose, however, which amino acids could be changed and still retain the claimed activity, and it did not disclose any actual variants of the defined amino acid sequence. *Id.* The examiner in *Kubin* rejected the claims as lacking enablement for sequences having identity to the defined amino acid sequence because of the absence of working examples and because changes in defined amino acid sequence might alter the function of the variant as compared to the defined amino acid sequence. *Id.* at 10. The examiner there also noted the unpredictability of the molecular biology art. *Id.* at 13. In finding enablement of the claimed invention, the Board agreed with the examiner that the molecular biology art was unpredictable (*Wands* factor 7), but "the other *Wands* factors weigh[ed] in

Appellants' favor, particularly the state of the art and the relative skill of those in the art as evidenced by the prior art teachings and Appellants' Specification." *Id.* at 14 (internal citations and markings omitted). Further, the Board noted that "[t]he amount of experimentation to practice the full scope of the claimed invention might have been extensive, but it would have been routine. The techniques necessary to do so were well known to those skilled in the art." *Id.* (emphasis added). Like the *Kubin* Appellants, Applicants here provided teachings on how to make variants of SEQ ID NO:18 (*see, e.g.,* Applicants' Specification at page 15, line 4 – page 16, line 22), described how to calculate the sequence identities between SEQ ID NO:18 and its variants (*see, e.g.,* Applicants' Specification at page 8, line 34 – page 9, line 17), and provided the Shannon assay to test for brittle-1 activity. Thus, *Kubin* dictates that Applicants' claims are enabled.

Applicants further note that, if Applicants' claimed invention is limited to only those nucleotide sequences encoding SEQ ID NO:18 as suggested by the Examiner, Applicants' patent rights become essentially useless because the skilled artisan could simply modify one amino acid of SEQ ID NO:18 (the sequence of which is undisputedly disclosed in Applicants' Specification), confirm brittle-1 activity by the Shannon assay (undisputedly referenced in Applicants' Specification), yet be outside the scope of the Applicants' claims even though Applicants' Specification disclosed the complete roadmap to working around the exceptionally narrow claims. In essence, the Examiner's scope of enablement rejection produces the absurd result of Applicants' Specification enabling the skilled artisan to avoid infringement of claims covering only nucleotide sequences encoding SEQ ID NO:18, but the same specification failing to enable the same skilled artisan to produce the same modified amino acid sequence if the claims cover sequences having 90% identity to SEQ ID NO:18.

Applicants also believe that any of the arguments presented in the enablement section should be applicable towards establishing that sufficient written description was present in Applicants' Specification as filed and vice versa. As noted in *LizardTech*, "a recitation of how to make and use the invention across the full breadth of the claim is ordinarily sufficient to demonstrate that the inventor possesses the full scope of the invention, and vice versa." 434 F.3d at 1345, 76

USPQ2d at 1733. That the present specification supports possession (written description) of the genus of polypeptides encompassed by the present claims (see above) further evidences enablement of the present claims. All methods for generating the described polypeptide variants were standard in the art at the time of filing. Likewise, methods for testing for the required activity were described in Applicants' Specification (see above). Thus, the possessed genus is enabled, almost by definition.

In view of the foregoing, Applicants respectfully request withdrawal of the Section 112, 1st paragraph, enablement rejections.

## VIII. CONCLUSION

For the reasons set forth above, the Board is respectfully requested to reverse the final rejection of pending Claims 26-29 and indicate allowability of all claims.

Please charge any fee due which is not accounted for to Deposit Account No. 501447 (Potter Anderson & Corroon LLP).

By: _____
Jeffrey Safran
Attorney for Applicants
Reg. No.: 54,689
Telephone: (302) 984-6132
Facsimile: (302) 658-1192

Dated: October 30, 2007

## CLAIMS APPENDIX

26. An isolated polynucleotide comprising:

(a) a nucleotide sequence encoding a polypeptide having brittle-1 activity, wherein the polypeptide has an amino acid sequence of at least 90% sequence identity when compared to SEQ ID NO:18, or

(b) a full-length complement of the nucleotide sequence of (a).

27. The isolated polynucleotide of Claim 26, wherein the polypeptide has a sequence identity of at least 95%, based on the Clustal method of alignment, when compared to SEQ ID NO:18.

28. A recombinant DNA construct comprising the polynucleotide of Claim 26 operably linked to a regulatory sequence.

29. A vector comprising the polynucleotide of Claim 26.

## EVIDENCE APPENDIX

A terminal disclaimer, attached herewith, was filed on December 21, 2005, disclaiming the terminal portion of any patent on this application that would extend beyond the expiration date of 6,660,850. The terminal declaimer was reviewed, accepted, and recorded.

Also attached herewith are the following references: Shannon *et al.*, Plant Physiol. 117:1235-52 (1998); You L. & Arnold F.H., Protein Eng. 9:77-83 (1996); Crameri A. *et al.*, Nat. Biotech. 15:436-38 (1997); Wu T.D. & Brutlag D.L., Proc. Int. Conf. Intell. Syst. Mol. Biol. 4:230-40 (1996); Cao H. & Shannon, J.C., Physiol. Plantarum 97:665-73 (1996); Shannon J.C. *et al.*, Plant Physiol. 110:835-43 (1996).

## RELATED PROCEEDINGS APPENDIX

None

Plant Physiol. (1998) 117: 1235–1252

# Brittle-1, an Adenylate Translocator, Facilitates Transfer of Extraplastidial Synthesized ADP-Glucose into Amyloplasts of Maize Endosperms[1]

**Jack C. Shannon*, Fang-Mei Pien, Heping Cao[2], and Kang-Chien Liu**

Department of Horticulture, 102 Tyson Building, The Pennsylvania State University, University Park, Pennsylvania 16802

**Amyloplasts of starchy tissues such as those of maize (*Zea mays* L.) function in the synthesis and accumulation of starch during kernel development. ADP-glucose pyrophosphorylase (AGPase) is known to be located in chloroplasts, and for many years it was generally accepted that AGPase was also localized in amyloplasts of starchy tissues. Recent aqueous fractionation of young maize endosperm led to the conclusion that 95% of the cellular AGPase was extraplastidial, but immunolocalization studies at the electron- and light-microscopic levels supported the conclusion that maize endosperm AGPase was localized in the amyloplasts. We report the results of two nonaqueous procedures that provide evidence that in maize endosperms in the linear phase of starch accumulation, 90% or more of the cellular AGPase is extraplastidial. We also provide evidence that the brittle-1 protein (BT1), an adenylate translocator with a KTGGL motif common to the ADP-glucose-binding site of starch synthases and bacterial glycogen synthases, functions in the transfer of ADP-glucose into the amyloplast stroma. The importance of the BT1 translocator in starch accumulation in maize endosperms is demonstrated by the severely reduced starch content in *bt1* mutant kernels.**

Starch is synthesized and accumulates in the amyloplasts of storage tissues (Shannon and Garwood, 1984; Boyer et al., 1989; Smith et al., 1997). The enzymatic reactions catalyzed by AGPases (EC 2.7.7.27), starch synthases (EC 2.4.1.21) (Preiss, 1991), SBEs (EC 2.4.1.18) (Cao and Preiss, 1996; Fisher et al., 1996), and starch-debranching enzymes (James et al., 1995; Rahman et al., 1998) are much better understood than the mechanism involved in the transport of substrates across the amyloplast envelope membranes and the compartmentation of AGPase (Pozueta-Romero et al., 1991; Liu et al., 1992; Okita, 1992; Hannah et al., 1993; Villand and Kleczkowski, 1994; Denyer et al., 1996; Pien and Shannon, 1996; Shannon et al., 1996; Thorbjornsen et al., 1996; Möhlmann et al., 1997). One of the major factors hindering progress is the difficulty of isolating highly purified intact amyloplasts and amyloplast membranes from storage organs because of the presence of a dense starch granule(s) within the fragile envelope membrane (Liu and Shannon, 1981; Echeverria et al., 1985; Gardner et al., 1987; Shannon et al., 1987; Shannon 1989).

We recently developed a rapid yet gentle procedure for the isolation of intact amyloplasts and their envelope membranes from immature maize (*Zea mays* L.) endosperms (Cao et al., 1995) and from maize endosperm suspension cultures (Cao and Shannon, 1996). Immunological characterization indicated that *Bt1* encodes the major 39- to 44-kD polypeptides of the purified amyloplast membranes, BT1. Results from several studies support the possibility that BT1 plays a significant role in starch accumulation in maize endosperm. For example, BT1 is specifically deficient in the amyloplast envelope membranes isolated from *bt1*, a starch-deficient endosperm mutant (Cao et al., 1995).

Shannon et al. (1996) demonstrated that ADP-Glc, the direct substrate for starch synthesis, accumulated in *bt1* mutant endosperms and that AGPase is the predominant enzyme responsible for the synthesis of ADP-Glc in *bt1*. In a preliminary report we showed that amyloplasts from young kernels isolated from *bt1* endosperms were only 25% as active in ADP-Glc uptake and conversion to starch as amyloplasts from normal and mutant maize endosperms (Liu et al., 1992). The amino acid sequence deduced from *Bt1* cDNA (Sullivan et al., 1991) shows high homology with mitochondrial adenylate translocators from some species, and in vitro-synthesized BT1 is targeted to the inner chloroplast membrane (Li et al., 1992).

Giroux and Hannah (1994) reported that the BT2 and SH2 subunits of AGPase from maize endosperms are the same size as the recombinant subunits, and suggested that AGPase in maize endosperm may not be located in amyloplasts. Denyer et al. (1996) recently provided evidence that maize endosperm cells contain two isozymes of AGPase, with more than 95% of the total activity being extra-amyloplastic. All of these data support the sugges-

[1] This work was supported by U.S. Department of Agriculture Competitive Grant no. 94-37306-0737.

[2] Present address: Department of Biochemistry and Biophysics, 2154 Molecular Biology Building, Iowa State University, Ames, IA 50011.

* Corresponding author; e-mail jshannon@psu.edu; fax 1–814–863–6139.

Abbreviations: ADH, alcohol dehydrogenase; AGPase, ADP-Glc pyrophosphorylase; APase, alkaline pyrophosphatase; BT1 and BT2, brittle-1 and brittle-2 proteins, respectively; DPP, days postpollination; FSBA, 5′-*p*-fluorosulfonylbenzoyl adenosine; hexose-P, hexose phosphate; 3-PGA, 3-phosphoglyceraldehyde; SBE, soluble starch-branching enzyme; SH2, shrunken-2 protein; SS, soluble starch synthase; SUS, Suc synthase; TCE, tetrachloroethylene; UGPase, UDP-Glc pyrophosphorylase.

tion that most of the ADP-Glc required for starch synthesis in maize endosperm is synthesized by cytosolic AGPase (Denyer et al., 1996), and that BT1 is the adenylate translocator responsible for the transfer of ADP-Glc into maize endosperm amyloplasts (Cao et al., 1995; Sullivan and Kaneko, 1995). Homologs of BT1 may be present in the amyloplast membranes from other starchy tissues, but they are not recognized by the antibodies to BT1 used by Cao et al. (1995) and by Cao and Shannon (1996, 1997).

Cytosolic localization of AGPase in maize endosperm cells is not supported by recent immunolocalization studies. For example, in a study using a transmission electron microscope, Miller and Chourey (1995) reported that proteins recognized by antibodies to spinach leaf AGPase were confined to amyloplasts, whereas antibodies to the peptide subunits of maize endosperm AGPase, BT2 and SH2, most heavily immunolabeled the amyloplasts and cell walls, with lighter labeling of the cytosol. In an in situ immunolocalization study at the light-microscopic level, Brangeon et al. (1997) observed that BT2 and SH2 antibodies (the same source of antibodies used by Miller and Chourey [1995]) immunolabeled both the amyloplasts and surrounding cytosol in pericarp cells from very young kernels, but immunolabel in endosperm cells from older kernels was closely associated with the amyloplasts only. These authors concluded that AGPase was localized in the amyloplast stroma of endosperm cells. However, at this level of resolution it is not possible to determine conclusively whether the immunolabeled proteins are in the plastid stroma or outside the envelope, and they correctly noted that the AGPase could have been bound to the outer membrane of the plastid envelope (Brangeon et al., 1997), and thus would partition as a "cytosolic" enzyme during aqueous fractionation.

A potential drawback of immunocytolocalization studies of cereal endosperm tissues at the electron-microscopic level is the difficulty encountered in sufficiently embedding the tissues so that the thin slices of starch granules do not "pop" out of the plastic before viewing. As a consequence, only amyloplasts with very small starch granules in cells located in the physiologically less-developed parts of the endosperm survive preparation for electron-microscopic examination. The surviving sections may or may not be representative of the entire tissue. Although this difficulty is minimized by using thicker sections for immunolocalization at the light-microscopic level, the resolution is not adequate to distinguish protein localization inside or outside of the plastid membranes.

There are also drawbacks to studies of enzyme compartmentation based on aqueously isolated amyloplasts. For example, during aqueous isolation most of the amyloplasts with starch granules larger than 1 or 2 $\mu$m in diameter are ruptured and the resulting preparation is enriched with amyloplasts containing smaller starch granules. To obtain the highest yield of intact amyloplasts, endosperms from very young kernels just beginning starch accumulation are used (Shannon et al., 1987). Activities of AGPase and SS were very low or undetectable in endosperms from 12-DPP kernels (Tsai et al., 1970), and Brangeon et al. (1997) showed a gradient of expression of the genes encoding AGPase from the periphery of the endosperm toward the center, with central endosperm cells of kernels 15 DPP most intensely immunolabeled by antibodies to BT2 and SH2. As a consequence, enzyme compartmentation in amyloplasts from very young kernels or from the physiologically younger cells near the periphery of the endosperm may not be representative of compartmentation in amyloplasts from those cells most actively engaged in starch biosynthesis.

To overcome these difficulties we developed two nonaqueous fractionation procedures to determine compartmentation of enzymes in amyloplasts from maize kernels in the linear phase of starch accumulation (about 20 DPP). Results of these studies are compared with results of an aqueous subcellular fractionation/immunoblotting study. Finally, results of a study of the uptake and incorporation into starch of metabolites by intact amyloplasts isolated from normal and mutant endosperms are reported. These studies support the conclusion that maize endosperm cells contain an extraplastidial form of AGPase, and that the amyloplast membrane-specific polypeptide, BT1, is an adenylate translocator.

## MATERIALS AND METHODS

Plant materials were either grown in the field at the Russell E. Larson Agricultural Research Farm (Centre County, PA) or grown in 20-L plastic pots containing two parts peat, two parts perlite, and one part soil. Potted plants were grown in the greenhouse in late winter and spring or were started in the greenhouse in the spring and then transferred outside the greenhouse for continued growth. High-intensity sodium lamps were used in the greenhouse to extend the daylength to 16 h. Unless noted otherwise, the normal maize (*Zea mays* L.) inbred W64A and the endosperm mutant genotypes *waxy* (*wx*), *brittle-1* (*bt1*), and *shrunken-2* (*sh2*) in a near-isogenic W64A background were used in these studies.

### Nonaqueous Compartmentation Studies

#### *Estimating Amyloplast Compartmentation of Enzymes after Fractionation in Mixtures of TCE and Heptane*

Endosperms from 20-DPP W64A inbred kernels were removed, frozen in liquid nitrogen, and freeze-dried. Pulverized samples were sifted through a 20-$\mu$m sieve using a sonic sifter fitted with a horizontal pulse generator (ATM Corp., Milwaukee, WI). The TCE/heptane procedure was patterned after methods used by Riens et al. (1991) and MacDougall et al. (1995). A total of 400 mg of dry, sifted endosperm in 50-mg batches was homogenized in 15-mL polypropylene centrifuge tubes in 2 mL of dry TCE using an ultrasonic probe (Biosonic IIA, Bronwill Scientific, Rochester, NY) for a total of 2.5 min using 30-s bursts. During homogenization the tube was held in a 95% ethanol/dry ice bath to reduce heating of the sample. Molecular sieve beads (4 Å, Sigma) were added to all TCE and heptane solutions before use to remove all traces of water, and

care was taken to keep all tubes tightly closed whenever possible.

The TCE homogenates were combined and *n*-heptane added to give a TCE:heptane mixture of 85:15 (v/v). Aliquots of this mixture were removed for "total-homogenate" analyses. The balance of the mixture was dispensed into several microcentrifuge tubes and the cellular contents were fractionated into amyloplast- and cytosol-enriched fractions by differentially pelleting the starch granules and associated enzymes from TCE/heptane mixtures of varying density. For example, the most dense fraction was pelleted from the 85:15 (v/v) TCE:heptane suspension by centrifugation in the cold (4°C) for 5 min at 16,000*g*. The pellet (pellet A) was retained and the supernatant was diluted with heptane to a final TCE:heptane ratio of 83:17 (v/v). Centrifugation was then repeated to yield pellet B. The resulting supernatant was again diluted with heptane to a ratio of 75:25 (v/v) and centrifuged as before to yield pellet C.

The 75:25 (v/v) TCE:heptane supernatant yielded the cytosol-enriched fraction. Aliquots of the initial unfractionated homogenate and the final cytosol-enriched supernatant were diluted with 3 volumes of heptane, and the particulate material in these heptane-diluted samples was collected by centrifugation in the cold for 10 min at 3000*g*. The clear supernatants were discarded and all pellets were held overnight at 4°C in a vacuum desiccator containing paraffin oil and silica-gel desiccant to remove the residual TCE and heptane. The dried pellets were extracted for enzyme analysis and the number of starch granules was determined. The TCE/heptane fractionation was repeated three times.

Amyloplasts in W64A endosperms each contain a single starch granule (Liu and Shannon, 1981); therefore, starch granule number was used as a measure of amyloplast number in the unfractionated homogenates and in the TCE/heptane fractions. We determined that the enzyme activities per million starch granules in the two most dense TCE/heptane fractions (pellets A and B) were very similar, and thus the means of both fractions ± SE (six values) were plotted. Likewise, activities per million starch granules in the unfractionated homogenate and in aqueous extracts of the sifted endosperm samples were very similar, and the means ± SE (six values) of these were plotted. The data from pellet C and supernatant fractions are the means ± SE of the three fractionations. To estimate compartmentation of an enzyme in amyloplasts, the average enzyme activity per million starch granules (*y* axis) from the four fractions (pellets A and B, unfractionated homogenate/aqueous extract, pellet C, and the supernatant) was plotted against the activity of a cytosol or vacuole marker enzyme per million starch granules from the same fractions (*x* axis). The *y* intersect of a regression line gives an estimate of enzyme activity per million amyloplasts in the absence of cytosol or vacuole contamination.

### *Glycerol Isolation of Amyloplasts*

A procedure for the nonaqueous isolation of starch granules with associated metabolites (amyloplasts) from maize endosperm amyloplasts was reported previously (Liu and Shannon, 1981). In that procedure the dry endosperm sample was homogenized in dry glycerol and filtered through Miracloth (Calbiochem), and the starch granules were pelleted through a more dense solution of 3-Cl-1,2-propanediol. Although this procedure yielded an amyloplast fraction essentially free of nuclear and cytosolic contaminants, the starch biosynthetic enzymes were inactivated. We determined that inactivation was caused primarily by excessive heating of the sample during homogenization in glycerol and by exposure to 3-Cl-1,2-propanediol. The glycerol nonaqueous isolation procedure was therefore modified for the enzyme-compartmentation study. Fifty milligrams of sifted endosperm as used above was added to a microcentrifuge tube containing 1 mL of dry, cold (4°C) glycerol. The sample was thoroughly dispersed using a disposable plastic microtube pestle. The microcentrifuge tube was closed and placed on ice, and centrifugation was carried out at 4°C for 20 min at 25,000*g*. The supernatant was transferred to a 15-mL centrifuge tube and the pellet was washed with 0.5 mL of dry, cold glycerol by suspension using the microtube pestle and centrifugation as above. After the wash supernatant was added to the initial supernatant, the wall of the microcentrifuge tube was wiped with a tissue to remove excess glycerol. The combined supernatants and pellet were retained for enzyme analyses. Total enzyme activity in aqueous extracts of the sifted endosperm was also determined.

### *Enzyme Extraction and Assay*

Duplicate samples of the TCE/heptane homogenate and of the four fractions were retained for enzyme analysis. Each fraction retained for extraction and enzyme assay was derived from approximately 42 mg of sifted endosperm. For the controls, duplicate subsamples (50 mg) of the sifted endosperm were also extracted and enzyme activities determined. Soluble enzymes were extracted from all pellet fractions and the sifted endosperm samples (TCE/heptane and glycerol) with 2 mL of HSB extraction buffer (50 mM Hepes, pH 7.5, 0.5 M sorbitol, 10 mM KCl, 1 mM $MgCl_2 \cdot 6H_2O$, 1 mM EDTA, 5 mM dithioerythritol, and 0.1% BSA) by sonication for four 10-s bursts with 10-s rest periods between each burst using the Biosonic IIA ultrasonic probe set at 60% maximum power. The tubes were held in an ice bath during sonication.

The glycerol supernatants (approximately 1.5 mL) were diluted to 5 mL with the HSB extraction buffer. The homogenates and diluted glycerol supernatants were centrifuged in the cold (4°C) for 10 min at 3000*g* and the supernatants retained for enzyme assay. Extracts from one set of the TCE/heptane pellets were used for assay of AGPase, UGPase, and ADH. AGPase and UGPase were assayed by the coupled-spectrometric method as described by Oh-Lee and Setter (1985), except that the AGPase and UGPase assays were started by the addition of 0.4 mM ADP-Glc and 0.4 mM UDP-Glc, respectively, and ADH as described by Cao et al. (1995). A small number of fractions were extracted at a time and AGPase, the most labile enzyme of the three, was assayed first. Extracts from the second set of

TCE/heptane pellets were used for assay of α-mannosidase as described by Boller and Kende (1979), for assay of APase as described by Gross and ap Rees (1986), and for assay of SBE and SS as described by Shannon et al. (1996). SUS was assayed in the hydrolytic direction as described by Echeverria et al. (1988), except that after heat inactivation, the quantity of Fru released was determined by a reducing sugar test as described previously (Shannon et al., 1996). The data were corrected for any Fru released in the absence of added UDP. HSB extracts of the glycerol fractions were used for assay of ADH, UGPase, AGPase, SBE, SS, and APase as described above.

### *Starch Granule Number*

The number of starch granules remaining in the pellets after HSB extraction for enzyme assays was determined as described by Shannon et al. (1996).

## Aqueous Compartmentation Study

### *Purification of Amyloplasts and Separation of Amyloplast Membranes and Stroma*

Crude and Percoll-purified amyloplasts were isolated from the endosperm of freshly harvested developing kernels (13–16 DPP) as described by Cao et al. (1995). The purified amyloplast pellet was suspended in a small volume of TDEP buffer (10 mM Tricine, pH 7.2, 1 mM DTT, 1 mM EDTA, and 0.5 mM PMSF), and the amyloplasts were lysed by one cycle of freezing at −70°C and thawing at 30°C. After removal of starch granules by centrifugation at 800*g*, amyloplast stroma was separated from amyloplast membranes by centrifugation at 100,000*g* for 60 min. Amyloplast membranes were further purified from the crude membrane pellet through a discontinuous Suc-density gradient as described by Cao et al. (1995). The purified amyloplast membrane pellet was suspended in TDEP buffer plus 0.2 M Suc and stored at −70°C.

### *Isolation of Microsomal Membranes*

Freshly isolated endosperms and other tissues were homogenized in a buffer containing 0.4 M Suc, 50 mM Mops, pH 6.9, 10 mM DTT, 1 mM EDTA, 0.1 mM PMSF, and 0.1% (w/v) BSA, and the homogenate was fractionated by differential centrifugation at 2,000*g* (P2), 10,000*g* (P10), and 100,000*g* (P100) according to the method of Cao et al. (1995). The P2 and P10 pellets and the microsomal membrane pellet (P100) were suspended in TES buffer (10 mM Tricine, pH 7.2, 1 mM EDTA, and 0.2 M Suc) and stored at −70°C.

### *Marker-Enzyme Analysis*

SBE (an amyloplast marker) and ADH (a cytosol marker) were assayed as described above. Catalase (EC 1.11.1.6) (a marker for microbodies), Cyt *c* oxidase (EC 1.9.3.1) (a marker for mitochondria), cyanide-insensitive NADH-Cyt *c* reductase (EC 1.6.99.3) (a marker for the ER), and vanadate-sensitive ATPase (EC 3.6.1.4) (a marker for plasma membrane) were assayed as described by Cao et al. (1995). Potassium-stimulated ATPase (a marker for plasma membrane), Triton-stimulated UDPase (a marker for the Golgi), and nitrate-sensitive ATPase (a marker for the tonoplast) were assayed as described by Briskin et al. (1987). Protein contents were measured using the Bradford method plus NaOH treatment (Cao et al., 1995).

### *SDS-PAGE and Immunoblotting*

Proteins were solubilized and denatured in 1× SDS gel-loading buffer by heating the samples in a boiling-water bath for 5 min. Polypeptides were separated by SDS-PAGE and stained with Coomassie brilliant blue R-250 (Cao et al., 1995). Standard procedures were used for immunoblotting, as described previously (Cao et al., 1995). The polyclonal antibodies to maize SH2 and BT2 were gifts from Michael Giroux and L. Curtis Hannah, University of Florida (Giroux and Hannah, 1994), and the polyclonal antibodies to maize BT1 were a gift from Thomas D. Sullivan, University of Wisconsin-Madison (Sullivan and Kaneko, 1995). The relative quantities of BT1, SH2, and BT2 were determined by scanning densitometry (model 300B, Molecular Dynamics, Sunnyvale, CA) using a method similar to that described by Cao et al. (1995).

### *Calculation of the Cellular Localization of the BT2- and SH2-Antibody-Reacting Polypeptides of AGPase*

The following data and calculations were used to determine the percentages of the SH2 and BT2 polypeptides of AGPase in the cytosol and amyloplasts. Total protein in the homogenate and crude amyloplast fraction (see Table III) was 3480 and 290 mg $g^{-1}$ fresh weight, respectively. Thirty-one percent of the amyloplast marker enzyme (SBE) was recovered in the crude amyloplast fraction (amyloplast yield). The crude amyloplast fraction was contaminated with 0.7% cytosol (percentage of the cytosol marker enzyme ADH). Based on equal loading of proteins and as determined by immunoblotting analysis and densitometer scanning (see Fig. 4), we estimated that the crude amyloplast fraction contained 39% and 18% of the cellular SH2 and BT2 polypeptides, respectively.

To determine the percentage of SH2 compartmented in the cytosol, we set the homogenate (the crude amyloplast fraction [*A*] and the cytosol fraction [*C*]) containing 100% SH2 × 3480 mg of protein = 3480 SH2 units; and the SH2 in the crude amyloplast (*A* + 0.7% of *C*) fraction = 39% SH2 polypeptide × 290 mg of protein/31% (amyloplast yield) = 365 SH2 units. To solve for *C*: (*A* + *C*) − (*A* + 0.007*C*) = 3480 − 365; 0.993*C* = 3115; *C* = 3137 SH2 units and *A* = 3480 − 3137 = 343 SH2 units. Therefore, the percentage of SH2 in the cytosol = 3137/3480 × 100 = 90.1%. Based on a similar calculation we determined that the cytosol contained 95.8% of the cellular BT2 polypeptide, for an average of 93% of the cellular BT2- and SH2-antibody-reacting polypeptides of AGPase localized in the cytosol.

### Metabolite-Uptake Studies

#### *Radioactive Metabolites and Chemicals*

Radioactive Glc-1-P, Glc-6-P, and ADP-Glc, uniformly $^{14}C$ labeled in the carbohydrate moiety, were purchased from ICN. Substrates, cofactors, inhibitors, and enzymes were obtained from Sigma, and all other chemicals used were analytical reagent grade.

#### *Aqueous Amyloplast Isolation and Purification*

Endosperms were removed from kernels 10 to 16 DPP (the precise ages are given in the tables) and homogenized in approximately 1 volume (w/v) of homogenization buffer (50 mM Hepes, pH 7.5, 0.5 M sorbitol, 10 mM KCl , 1 mM $MgCl_2$, 1 mM EDTA, 0.1% BSA, and 5 mM dithioerythritol) for 2 s at top speed in a homogenizer (VirTis 23, The VirTis Co., Gardiner, NY). The homogenate was gently filtered through Miracloth and an aliquot layered on a gradient of 10%, 20%, and 40% Percoll in the homogenization medium. The gradient was centrifuged for 5 min at 200*g* and the amyloplasts settling in the 20% Percoll layer were removed and used for the uptake studies. An aliquot of each preparation was removed to determine amyloplast intactness by measuring SBE activity before and after lysis, as described previously (Shannon et al., 1987).

#### *Metabolite Uptake and Incorporation*

For uptake and incorporation of ADP-Glc, amyloplasts (60–80 μL) were added to a reaction mixture (200 μL final volume) containing 100 mM Bicine, pH 8.5, 0.5 M sorbitol, 12.5 mM EDTA, 10 mM GSH, 50 mM $KC_2H_3O_2$, and 4 mM [$^{14}C$]ADP-Glc (the specific activity varied from 64 to 300 cpm/nmol). To determine the effect of ATP, ADP, or AMP on [$^{14}C$]ADP-Glc uptake and incorporation, the individual nucleotides (7 mM) were added to the amyloplasts in the 20% Percoll isolation buffer and incubated on ice for 30 min before an aliquot of the amyloplast suspension was added to the uptake medium (final nucleotide concentration in the uptake medium was 2.2 mM). To determine whether a translocator with an adenosine-binding site functions in the uptake of [$^{14}C$]ADP-Glc, the amyloplasts were preincubated for 30 min at 30°C in the uptake mixture containing varying concentrations of FSBA before the addition of [$^{14}C$]ADP-Glc. In the FSBA study each uptake solution contained 2% DMSO, the solvent for FSBA.

To determine the uptake and incorporation of Glc-1-P and Glc-6-P, amyloplasts (60–80 μL) were added to a reaction mixture (200 μL final volume) containing 15 mM Hepes, pH 7.5, 0.5 M sorbitol, 10 mM $MgCl_2$, 12.4 or 0.5 mM 3-PGA, 0.08% BSA, 0.1 unit of inorganic pyrophosphatase, and 2 mM [$^{14}C$]Glc-1-P (about 200 cpm/nmol) or [$^{14}C$]Glc-6-P (about 200 cpm/nmol). ATP at 2 mM and rabbit liver glycogen at 1 mg per uptake reaction were added as indicated.

All uptake studies were completed with intact amyloplast preparations and with lysed amyloplast preparations. There were no differences in the results when the amyloplasts were lysed either by including 1% Triton X-100 in the uptake medium or by brief sonication (four times for 10 s each, with cool-down periods between) of the uptake medium containing amyloplasts before the addition of the $^{14}C$-metabolite. Unless noted otherwise the uptake reactions were carried out at 30°C and were terminated after 120 min by addition of 2 mL of 75% methanol containing 1% KCl. The alcohol-insoluble pellet was collected by centrifugation (2000*g*) in the cold for 10 min, and was washed twice with the methanol/KCl solution by suspension and centrifugation as above. The alcohol-washed pellets were then extracted three times with water by suspension and centrifugation as described above. The quantities of $^{14}C$ product in the water-soluble and -insoluble fractions were determined using a liquid-scintillation analyzer (Tri-Carb 1500, Packard Instrument Co., Downers Grove, IL).

An aliquot from each amyloplast isolation used for uptake studies was retained to determine the number of starch granules in each uptake reaction. Starch granule number was as determined previously (Shannon et al., 1996). Uptake and incorporation data are presented as the amount per million starch granules (amyloplasts). It is assumed that each amyloplast settling in the 20% Percoll layer contains one starch granule.

### Protein-Sequence Analysis

The protein sequences used in the analysis were obtained from the literature and searched from the database of the National Center for Biotechnology Information. The locations of amino acid residues indicated in the tables correspond to the translated full-length sequences instead of the "mature" sequences. The sequence analysis was conducted as described previously for the alignment of branching enzymes (Cao and Preiss, 1996).

### Transmission Electron Microscopy

Kernels were removed from 20-DPP *bt1* ears and small portions of endosperm were removed from the middle of the kernel and fixed for 4 h at room temperature in 4% glutaraldehyde in 100 mM cacodylate buffer, pH 7.0. The samples were postfixed for 1 h in 1% osmium tetroxide in the same buffer and then dehydrated through an ethanol series and embedded in the ultra-low-viscosity medium (VCD/HXSA) described by Oliveira et al. (1983). Silver to gold sections were cut using a diamond knife and a microtome (model III-8800, LKB, Bromma, Sweden) and examined by a transmission electron microscope (model 1200EXII, Jeol) either without additional staining or after staining with uranyl acetate and lead citrate.

## RESULTS

### Nonaqueous Fractionations

#### *TCE/Heptane Fractionation*

The percentage of recovery of starch granules and the activity of selected enzymes in the three fractions pelleting at various densities and the supernatant were all in excess of 80% of that in the original homogenate (Fig. 1). Forty-

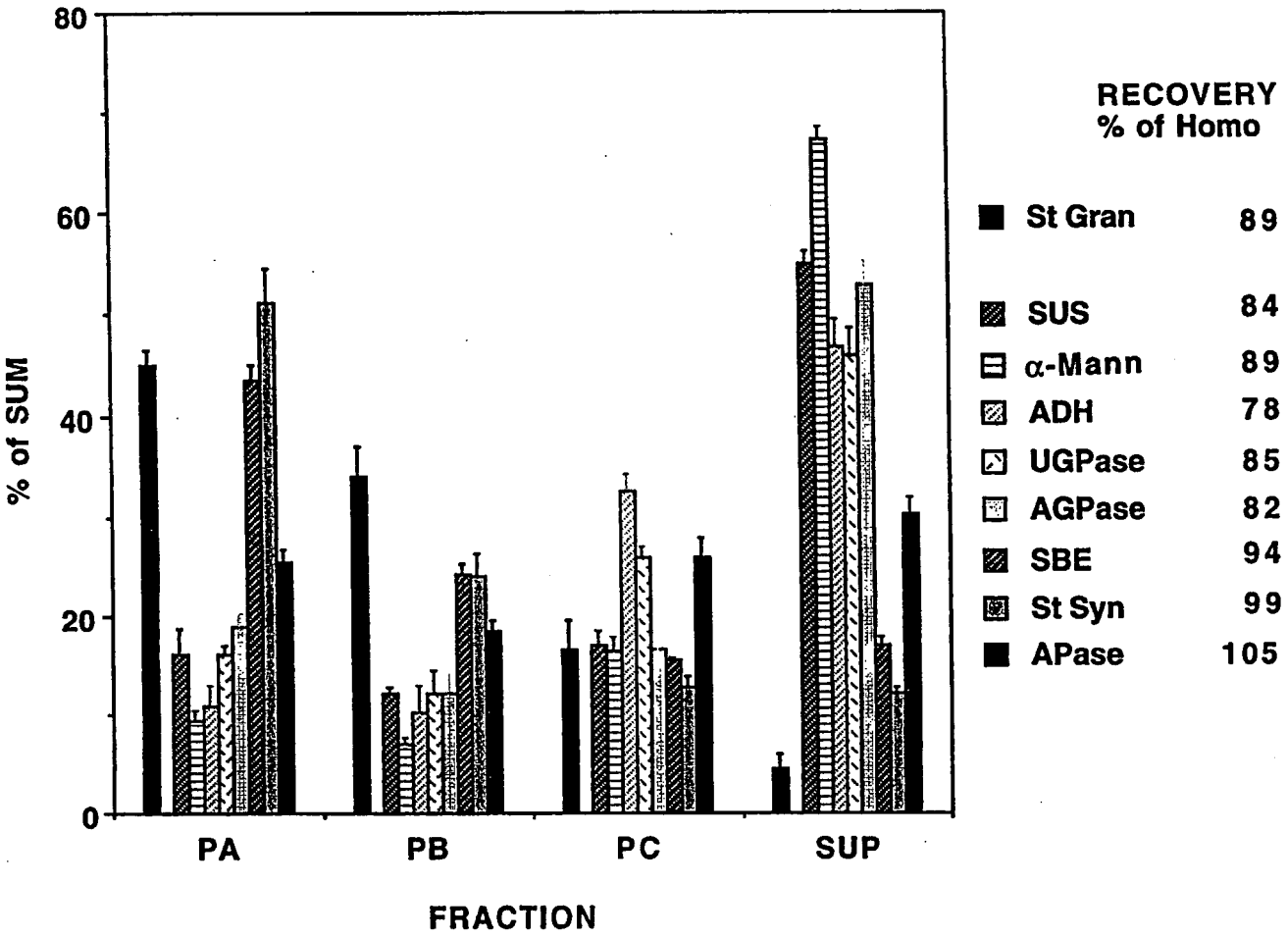



**Figure 1.** Distribution of starch granules and the activity of SUS, ADH, UGPase, AGPase, α-mannosidase (α-Mann), SBE, SS (St Syn), and APase in TCE/heptane fractions of different densities. The most dense fraction, pellet A (PA), was enriched in amyloplasts, and the least dense fraction, the supernatant (SUP) fraction, was enriched in cytosol. The distribution as a percentage of the sum of starch granules and enzyme activities from normal W64A endosperms and recovery of each as a percentage of the homogenate are recorded next to the figure key. Data are means ± SE of three fractionations. PB, Pellet B; PC, pellet C.

five percent of the starch granules was recovered in the most dense fraction (pellet A) and about 5% was recovered in the least dense fraction (the supernatant fraction). The amyloplast marker enzymes SBE and SS partitioned most closely with the starch granules, and the cytosol and vacuole marker enzymes, SUS, ADH, UGPase, and α-mannosidase, were low in the pellet A fraction and high in the supernatant fraction (Fig. 1). AGPase partitioned most closely with the cytosol marker enzymes and APase partitioned intermediate between the amyloplast and cytosol marker enzymes (Fig. 1).

The quantities of SBE, SS, APase, and AGPase associated with amyloplasts were determined from the *y* intercepts of the simple regression of the plot of target-enzyme activity per $10^6$ starch granules versus nonplastid marker-enzyme activity per $10^6$ starch granules. From this analysis we determined that 71%, 77%, and 58% of the putative amyloplast marker enzymes SBE, SS, and APase, respectively, were associated with the amyloplasts (Table I). The quantity of AGPase associated with the amyloplasts varied depending on the nonplastid marker used for the plot, but it is clear that little if any AGPase was recovered with the amyloplasts. These results are based on enzyme activities readily extracted in aqueous buffer solutions and are not expected to include the more tightly bound starch-granule-associated starch-synthase I and SBE II reported by Mu-Forster et al. (1996).

We intended to use the TCE/heptane procedure to estimate enzyme compartmentation in endosperm samples from the *bt1* mutant genotype. However, a critical part of the determination is an accurate count of the number of starch granules. We found that some of the TCE/heptane fractions contained a mixture of small starch granules (about 2 μm) and very small starch granules (less than 1 μm). Even though we stained the samples with iodine, we were unable to distinguish the smallest starch granules from protein bodies. In addition, the compartmentation calculation assumes that each amyloplast contains a single

**Table I.** *Compartmentation of enzymes in amyloplasts from W64A endosperm as determined using the TCE/heptane fractionation method*

SBE, SS, APase, and AGPase activities per million starch granules (*y* axis) in the TCE/heptane fractions were individually plotted against the activities per million starch granules of SUS, UGPase, ADH, and α-mannosidase (α-Mann), the nonplastidial marker enzymes. Estimates of SBE, SS, APase, and AGPase activities in amyloplasts per million starch granules were determined from the *y* intercept of a simple regression line from each individual plot. Activities of SBE, SS, APase, and AGPase per million starch granules in the total homogenate (Homo) are included. Data from three separate TCE/heptane fractionations were plotted.

| Nonplastidial Enzyme | Activity in Amyloplasts | | | |
|---|---|---|---|---|
| | SBE | SS | APase | AGPase |
| | *nmol min$^{-1}$ 10$^{-6}$ starch granules* | | | |
| SUS | 30.94 | 0.14 | 0.97 | 0.03 |
| UGPase | 27.80 | 0.14 | 0.65 | −1.39 |
| ADH | 26.83 | 0.14 | 0.53 | −1.84 |
| α-Mann | 33.17 | 0.15 | 1.17 | 1.00 |
| Mean ± SE | 29.68 ± 2.52 | 0.14 ± 0.01 | 0.83 ± 0.25 | −0.55 |
| | *nmol min$^{-1}$ 10$^{-6}$ starch granules* | | | |
| Homo activity | 41.91 ± 2.20 | 0.19 ± 0.04 | 1.42 ± 0.06 | 4.15 ± 0.70 |
| | *% of Homo* | | | |
| Amyloplast activity | 70.8 | 76.9 | 58.4 | 0.0 |

starch granule. This is true for the normal inbred W64A and all maize endosperm mutant genotypes in the W64A background examined to date except for *su1* (Shannon and Garwood, 1984). However, the presence of the very small starch granules in the *bt1* samples caused us to question this assumption and we prepared fresh *bt1* endosperm samples for transmission electron microscopic examination. From this examination it is clear that endosperm cells from 20-DPP *bt1* kernels contain two populations of amyloplasts: simple amyloplasts with a single starch granule 1 to 5 μm in diameter, and compound amyloplasts, containing several very small starch granules (1 μm or less) (Fig. 2). Therefore, we were unable to accurately estimate enzyme compartmentation in *bt1* endosperm cells by the TCE/heptane procedure.

### *Nonaqueous Glycerol Isolation*

The recovery of enzyme activity after the nonaqueous glycerol fractionation procedure varied between 40% and 117% of the enzyme activities measured after extraction in the HSB buffer (Table II). The glycerol-isolated amyloplast pellets contained only 8% and 7% of the cytosol marker enzymes, ADH and UGPase, respectively, and 14% of the AGPase activity (Table II). Thus, if we assume that the glycerol-isolated amyloplast pellet contains 7% cytosol contamination, then 7% of the cellular AGPase was compartmented in the amyloplasts. In contrast, 95%, 79%, and 38% of the recovered activities of the amyloplast marker enzymes SBE, SS, and APase partitioned with the glycerol-isolated amyloplasts, respectively. It is important to note that although the sum of APase activities in the glycerol supernatant and pellet fractions was 17% higher than in the HSB extract, only 56%, 40%, and 60% of the HSB-extractable activities of AGPase, SBE, and SS, respectively, were recovered in the glycerol-supernatant-plus-pellet fractions. The aliquots of HSB-diluted glycerol supernatant needed for assay of ADH, UGPase, and AGPase were smaller than those needed for assay of the plastid enzymes SBE and SS.

In a separate study we determined that the 0.3% to 1.5% of glycerol carried over from the diluted glycerol supernatant into the reaction mixtures was not inhibitory to ADH, UGPase, and AGPase, but that the 6% glycerol carried over into the assay mixtures for SBE and SS reduced measurable SBE and SS activities by approximately 60% and 25%, respectively, compared with assays in the absence of glycerol (data not shown). Thus, glycerol inhibition in the glycerol supernatants may contribute to the low recoveries of SBE and SS activities and inflate apparent partitioning of these enzymes in the glycerol-pellet fraction. When the percentage recoveries of the amyloplast marker enzymes associated with the glycerol pellets were calculated as percentages of the activity in the HSB extracts, we estimated that the glycerol pellets contained 38%, 47%, and 46% of the total cellular SBE, SS, and APase, respectively.

Activities of AGPase, SBE, SS, and APase per million starch granules in the glycerol-isolated amyloplasts (Table II) were very similar to the enzyme activities associated with the amyloplasts, as estimated by the TCE/heptane-fractionation procedure (Table I). It is clear from the results of these studies that a much higher percentage of AGPase partitions in the cytosol fraction compared with the amyloplast marker enzymes. In addition, we have shown that the nonaqueous glycerol procedure may be used to isolate starch-granule preparations from kernels in mid development (20 DPP), which contain almost half of the soluble stromal enzymes but are relatively free of cytosol marker enzymes.

## Aqueous Fractionation and Immunolocalization

### *Preparation of Amyloplasts from Developing Maize Endosperm*

As a second approach to confirm the subcellular localization of AGPase in maize endosperm, we isolated intact amyloplasts from 13-DPP endosperm. Typical examples of

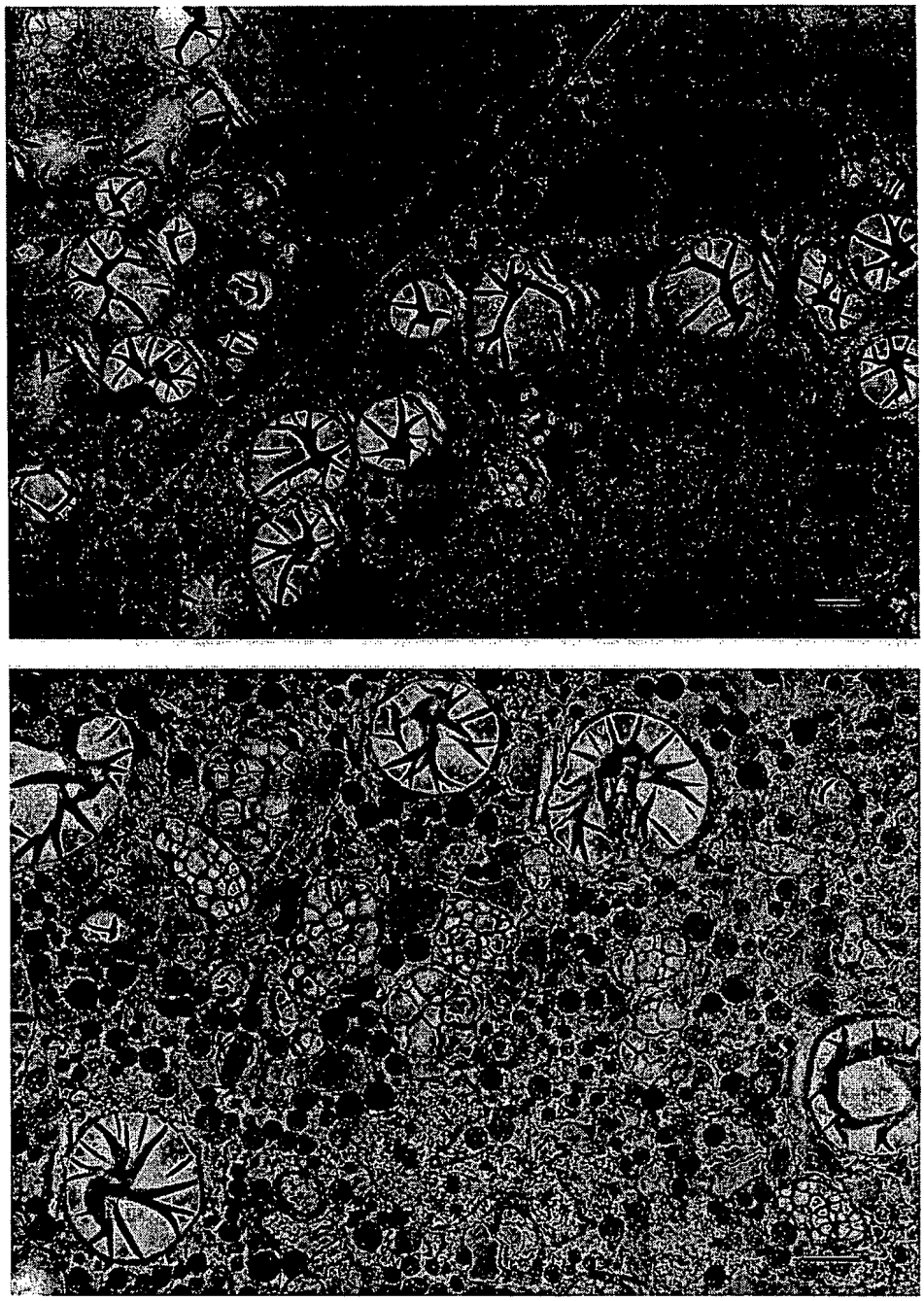


**Figure 2.** Low-magnification transmission electron photomicrographs of the parts of three cells in endosperm from a 20-DPP *bt1* mutant kernel. The section in the top micrograph was not poststained and that in the bottom micrograph was poststained with uranyl acetate and lead citrate. The large starch granules in the simple amyloplasts show dark artifacts that formed during sectioning because of the hydration and folding of the thin slices of starch. The compound amyloplasts contain many small starch granules. Bars = 2 $\mu$m.

the yield and purity of the aqueously isolated amyloplasts are summarized in Table III. Partially purified amyloplasts recovered in the 100*g* pellet (crude amyloplasts) contained approximately 31% of the amyloplast marker enzyme SBE and less than 1%, 4%, and 2% of the cytosol marker (ADH), the mitochondrial marker (Cyt *c* oxidase), and the ER marker (cyanide-insensitive NADH-Cyt *c* reductase) enzymes, respectively. Purification of the amyloplasts through a Percoll density gradient effectively removed all of the cytosol, mitochondria, and ER contaminants, but only 13.9% of the extractable SBE was retained with the more highly purified amyloplasts. Therefore, during Per-

**Table II.** *Partitioning of cytosol and amyloplast marker enzymes and AGPase in glycerol supernatant and pellet fractions*

Freeze-dried 20-DPP W64A endosperm tissues were pulverized and sifted through a 20-μm sieve. Samples of sifted endosperm were homogenized in glycerol (Gly) and separated into supernatant and pellet fractions. The activities in these fractions were compared with the total activities in subsamples of the sifted endosperm extracted in HSB buffer. Activities are presented per 50 mg of sifted endosperm sample and per million starch granules. Data are the average ± SE of the number of fractionations (shown in parentheses).

| Enzyme | Fractionation | Enzyme Activity | | | Recovery after Gly | Activity in Gly Pellet | Activity in HSB or Gly Pellets |
|---|---|---|---|---|---|---|---|
| | | Supernatant | Pellet | Total/sum | | | |
| | | *nmol min$^{-1}$ 50 mg$^{-1}$ dry wt* | | | *% of HSB* | *% of sum* | *nmol min$^{-1}$ 10$^{-6}$ starch granules* |
| ADH (5) | HSB | – | – | 1084 ± 202 | – | – | 9.77 |
| | Gly-HSB | 756 ± 80 | 67 ± 33 | 823 ± 93 | 76.0 | 8.2 ± 3.4 | 0.66 |
| UGPase (4) | HSB | – | – | 7935 ± 1860 | – | – | 71.49 |
| | Gly-HSB | 6457 ± 1009 | 483 ± 171 | 6940 ± 1055 | 87.5 | 7.0 ± 2.5 | 4.76 |
| AGPase (5) | HSB | – | – | 649.5 ± 108.7 | – | – | 5.85 |
| | Gly-HSB | 309.4 ± 72.4 | 51.2 ± 17.9 | 360.6 ± 88.3 | 55.5 | 14.2 ± 1.9 | 0.50 |
| SBE (4) | HSB | – | – | 6716 ± 920 | – | – | 60.51 |
| | Gly-HSB | 130 ± 59 | 2571 ± 187 | 2701 ± 228 | 40.2 | 95.2 ± 1.8 | 25.33 |
| SS (4) | HSB | – | – | 36.3 ± 6.5 | – | – | 0.33 |
| | Gly-HSB | 4.5 ± 0.6 | 17.2 ± 2.3 | 21.7 ± 2.9 | 59.8 | 79.3 ± 1.1 | 0.17 |
| APase (6) | HSB | – | – | 221.7 ± 39.2 | – | – | 1.99 |
| | Gly-HSB | 162.0 ± 32.3 | 98.4 ± 16.5 | 260.4 ± 44.3 | 117.5 | 37.9 ± 3.8 | 1.01 |

coll purification many of the amyloplasts were ruptured, releasing SBE from the amyloplasts. The crude and purified amyloplast preparations contained 8.3% and 3.4% of the homogenate protein, respectively, and the specific activity of SBE relative to that in the homogenate was enriched 3.7- and 4.1-fold in the crude and Percoll-purified amyloplast fractions, respectively (Table III).

Other cellular components such as the microbodies, plasma membrane, Golgi, and tonoplast cosedimented with the crude amyloplast preparation, resulting in a 2.2- to 5.5-fold increase in specific activity of these marker enzymes. However, after Percoll purification the amyloplast fraction was essentially free of catalase, the microbody marker, and contained only 0.6% of the cellular vanadate-sensitive ATPase, one of the markers for plasma membranes (Table III). The purified amyloplasts contained 5.9% of the cellular potassium-stimulated ATPase, a second putative plasma-membrane marker, but the specific activity had declined from 0.1 to 0.04. Likewise, the percentages of Triton-stimulated UDPase, a Golgi marker, and nitrate-sensitive ATPase, a tonoplast marker, were reduced to 4.2% and 3.5%, respectively, and their specific activities were much lower than in the crude pellet (Table III).

Membranes isolated from Percoll-purified amyloplasts were very yellow, with an absorption spectrum characteristic of carotenoids (plastid membrane marker): absorption peaks at 458 and 488 nm (data not shown). This membrane fraction was much enriched in the amyloplast membrane-specific polypeptide BT1 (Cao et al., 1995) compared with the total microsomal membranes (Fig. 3).

**Table III.** *Yield and purity of amyloplasts isolated from developing maize endosperm*

Amyloplasts were isolated from 13-DPP endosperm from cv Pioneer 3780 kernels. Aliquots were assayed for protein and marker enzymes after filtration through Miracloth (Homogenate), the first 100*g* centrifugation pellet (Crude Amyloplasts), and the 100*g* Percoll density-gradient-centrifugation pellet (Purified Amyloplasts). All samples were suspended in homogenization buffer and lysed by one freeze-and-thaw cycle before the starch was removed by centrifugation and the soluble protein content and marker enzyme activities were determined. Data are the average of two or three determinations. Homogenate protein is presented as milligrams per gram fresh weight and all enzyme results are presented as nanomoles per minute per gram fresh weight. Enzyme activities in the crude and purified amyloplasts are presented as a percentage of the activity in the homogenate (% of Homo) and as specific activity (Spec Act) nanomoles per minute per milligram protein.

| Marker Enzyme | Compartment | Homogenate | | Crude Amyloplasts | | Purified Amyloplasts | |
|---|---|---|---|---|---|---|---|
| | | Activity | Spec Act | % of Homo | Spec Act | % of Homo | Spec Act |
| Protein | – | 3,480 | – | 8.3 | – | 3.40 | – |
| SBE | Amyloplast | 1,800 | 0.517 | 31.1 | 1.931 | 13.90 | 2.137 |
| ADH | Cytosol | 3,913 | 1.124 | 0.7 | 0.090 | 0 | 0 |
| Cyt *c* oxidase | Mitochondria | 69 | 0.020 | 3.6 | 0.009 | 0 | 0 |
| NADH Cyt *c* reductase[a] | ER | 1,113 | 0.320 | 1.7 | 0.064 | 0.02 | 0.002 |
| Catalase | Microbodies | 81,300 | 23.362 | 18.5 | 51.724 | 0.04 | 0.256 |
| Vanadate-sensitive ATPase | Plasma membrane | 143 | 0.041 | 18.3 | 0.090 | 0.60 | 0.008 |
| Potassium-stimulated ATPase | Plasma membrane | 74 | 0.021 | 39.9 | 0.102 | 5.90 | 0.038 |
| Triton X-100-stimulated UDPase[b] | Golgi | 121 | 0.035 | 46.6 | 0.194 | 4.20 | 0.044 |
| Nitrate-sensitive ATPase | Tonoplast | 76 | 0.022 | 33.6 | 0.088 | 3.50 | 0.023 |

[a] Cyanide-insensitive activity. [b] Color reagent without 1.5% SDS was used.

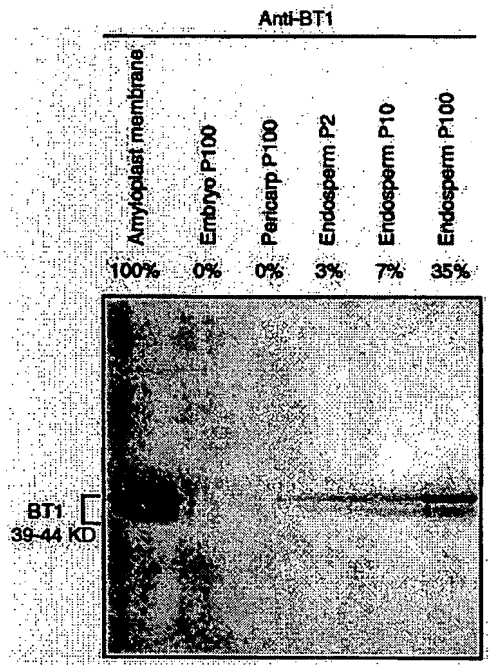



**Figure 3.** Immunolocalization of BT1 polypeptides in maize amyloplast membranes (lane 1), in microsomal membranes (P100) from the embryo, pericarp, and endosperm tissues (lanes 2, 3, and 6, respectively), and in pellets forming at 2,000*g* (P2) and 10,000*g* (P10) (lanes 4 and 5, respectively). The amyloplast membranes were isolated from endosperm amyloplasts purified from immature (approximately 12–15 DPP) Doebler 66XP hybrid kernels, and the P2, P10, and P100 fractions were isolated from 13-DPP Pioneer 3780 hybrid kernels as described previously (Cao et al., 1995). Polypeptides were separated by SDS-PAGE (15% separating gel), transferred to a nitrocellulose filter, and probed with polyclonal antibodies raised against a fusion protein containing 56 amino acids of the C terminus of BT1 and glutathione *S*-transferase (Sullivan and Kaneko, 1995). Lane 1 contained 8 μg of amyloplast-membrane protein; all other lanes contained 30 μg of protein. The relative quantities of BT1 in the various lanes as shown on the figure were determined by densitometry.

### *Immunolocalization of the BT2- and SH2-Antibody-Reacting Polypeptides of AGPase*

Because the enrichment of amyloplasts based on the specific activity of SBE in the crude and Percoll-purified amyloplast preparations was similar but the yield of the amyloplast marker in crude amyloplasts was much higher than that in the purified amyloplast preparation, we chose crude amyloplasts for this experiment. Proteins from endosperm homogenates and crude amyloplasts were separated by 15% separating gel, transferred to nitrocellulose membranes, and probed with polyclonal antibodies raised against maize SH2 and BT2 polypeptides. The same-size polypeptides were detected in both the whole homogenate and the crude amyloplast preparations (Fig. 4).

The most significant result was that when equal quantities of protein were loaded, the intensities of SH2 or BT2 antibody-reacting polypeptides(s) were not enriched in the proteins from the crude amyloplast fraction relative to those in the homogenate (Fig. 4). Rather, based on densitometer analyses we estimated that the levels of SH2 and BT2 polypeptides in the crude amyloplasts were about one-third and one-fifth of those in the homogenate, respectively (Fig. 4). This lack of enrichment in the crude amyloplasts of the AGPase polypeptides was in sharp contrast to the approximately 4-fold enrichment of extractable SBE, the amyloplast stroma marker enzyme (Table III), and the 10-fold enrichment of the amyloplast membrane marker BT1 in amyloplast membranes recovered from the Percoll-purified amyloplasts (Fig. 3). This lack of BT2 and SH2 enrichment clearly indicates that the majority of the BT2- and SH2-antibody-reacting AGPase was localized outside of the amyloplasts.

Because the yield of amyloplasts in the crude amyloplast preparation was 31%, based on SBE activity, and cytosol contamination was 0.7%, based on ADH activity, we estimated that 90% and 95.8% of the total SH2 and BT2 proteins, respectively, or an average of 93% of the cellular BT2- and SH2-antibody-reacting polypeptides of AGPase, were located in the cytosol (see "Materials and Methods" for calculations). These values are very close to the estimates of AGPase compartmentation determined by the TCE/heptane and glycerol-isolation procedures reported above (Tables I and II).

## Metabolite Uptake and Incorporation into Starch by Isolated Amyloplasts

### *Hexose-P Uptake and Incorporation*

Intact amyloplasts were aqueously isolated and purified from normal and mutant endosperms and their capacities

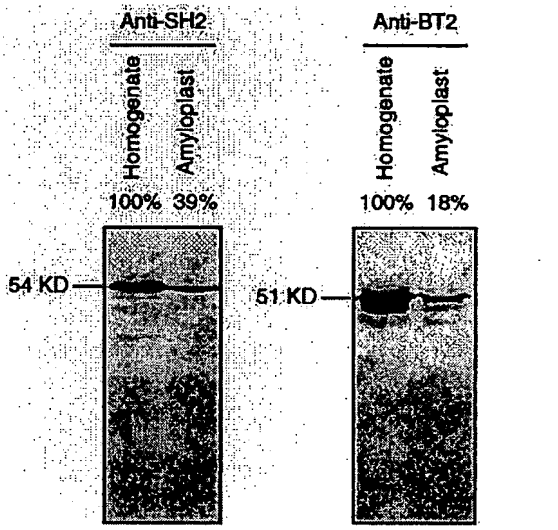



**Figure 4.** Immunolocalization of AGPase in the homogenate and in the crude amyloplasts (100*g* pellet) from maize endosperm cells. Polypeptides were separated by SDS-PAGE (15% separating gel), transferred onto nitrocellulose membranes, and probed with polyclonal antibodies raised against maize SH2 polypeptide (left) and maize BT2 polypeptide (right) (Giroux and Hannah, 1994). All lanes contained 12 μg of protein. Lanes 1 and 2 contained peptides from the homogenate and the crude amyloplasts, respectively. The relative quantity of SH2 and BT2 protein shown on the figure was determined by densitometry.

for the uptake and use of hexose-Ps and ADP-Glc for starch synthesis were determined. Sixty percent or more of the amyloplasts used for these studies were judged to be intact based on latency analysis (data not shown). Table IV summarizes the incorporation of [$^{14}$C]Glc into a methanol-insoluble product after incubation of intact or lysed amyloplasts in uptake medium containing [$^{14}$C]Glc-1-P or [$^{14}$C]Glc-6-P either with or without added ATP. In the absence of added glycogen, very little [$^{14}$C]Glc from either hexose-P was incorporated into the methanol-insoluble product. Generally, Glc transfer from Glc-1-P was somewhat higher than that from Glc-6-P, but there was little if any effect of added ATP, a substrate for AGPase. In addition, incorporation by amyloplast from the AGPase-deficient mutant *sh2* was equal to that by amyloplasts from normal, *wx* and *bt1* endosperms.

The similarity in incorporation of [$^{14}$C]Glc-1-P by all genotypes tested and the small effect of added ATP indicates that polymerization was most likely caused by the activity of starch phosphorylase, with little contribution by plastid-localized AGPase. The standard uptake solution used in these studies contained relatively high levels of 3-PGA (12.4 mM), the allosteric activator of AGPase. It is possible that the low incorporation of hexose-Ps was the result of 3-PGA inhibition of hexose-P uptake, but this was ruled out by a later study in which we showed that uptake and incorporation of hexose-Ps in the presence of 0.5 and 12.4 mM 3-PGA were similar (Table IV).

During incubation some of the amyloplasts are invariably ruptured, releasing plastid enzymes. The consequences of this were seen when glycogen, an alternative glucan acceptor, was added to the "intact" and lysed amyloplasts incubated with Glc-1-P. Over 10 times more [$^{14}$C]Glc was incorporated into the methanol-insoluble polymer but, again, there was little if any effect of added ATP (Table IV). This supports the conclusion that the isolated amyloplasts contain an active starch phosphorylase that effectively transfers Glc from Glc-1-P to glycogen, but was much less effective in Glc transfer to the native glucan acceptors (starch granules) of the amyloplasts. Apparently, amyloplasts isolated from *sh2* (the only genotype tested) contain very little phosphoglucomutase, or that which is present is essentially inactive in the uptake conditions used, because even in the presence of added glycogen there was very little transfer of [$^{14}$C]Glc from Glc-6-P to the methanol-insoluble product (Table IV).

### *ADP-Glc Uptake and Incorporation*

Intact maize amyloplasts isolated from 10- to 16-DPP normal, *wx*, and *sh2* endosperms incorporated more than 10 times as much Glc from ADP-Glc into a methanol- and water-insoluble product (Table V) as from Glc-1-P (Table IV). This difference in incorporation was not caused by a difference in the buffer salt or pH of the uptake solutions used in the standard conditions, because in a later study we determined that there was no difference in Glc incorporation from ADP-Glc when the uptake solution was buffered with Hepes at pH 7.5 rather than at pH 8.5 (Table V). Hydrolysis of the water-insoluble product with β-amylase yielded maltose and the product was judged to be starch (data not shown). Lysis of the amyloplasts before incubation reduced incorporation 70% to 90%. Among the three genotypes, intact amyloplasts from *sh2* most effectively converted ADP-Glc to starch. In contrast, intact amyloplasts from the other starch-deficient/high-sugar genotype, *bt1*, was only 26% as effective in the uptake and conversion of ADP-Glc to starch as amyloplasts from *wx*,

**Table IV.** *Hexose-P uptake and incorporation into methanol-insoluble product*

Amyloplasts were from the 20% Percoll fraction. The reaction mixture (200 μL final volume) contained 15 mM Hepes, pH 7.5, 0.08% BSA, 10 mM $MgCl_2$, 12.4 mM 3-PGA, 0.5 M sorbitol, 0.1 unit of inorganic pyrophosphatase, 2 mM Glc-1-P or Glc-6-P, 2 mM ATP, and 1 mg of rabbit-liver glycogen (RLG) as indicated plus 60 or 80 μL of amyloplast fraction. For the uptake studies amyloplasts were isolated from the endosperms of the following age kernels: N, 12 DPP; *wx*, two 12 DPP, one 13 DPP, and one 16 DPP for the minus-RLG study, and one 13 DPP and one 16 DPP for the plus-RLG study; *sh2*, one each of 13, 15, and 16 DPP; and *bt1*, one 11 DPP.

| Genotype (*n*) | Total Incorporation | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Glc-1-P | | | | Glc-6-P | | | |
| | Intact | | Lysed | | Intact | | Lysed | |
| | +ATP | −ATP | +ATP | −ATP | +ATP | −ATP | +ATP | −ATP |
| | *pmol min$^{-1}$ 10$^{-6}$ amyloplasts* | | | | | | | |
| No RLG | | | | | | | | |
| N (1) | 4.17 | 1.83 | 2.08 | – | – | – | – | – |
| *wx* (4) | 4.54 ± 0.50 | 3.90 ± 1.42 | 4.12 ± 1.22 | – | 1.96 ± 1.37 | 1.25 ± 0.42 | 1.08 | – |
| *wx* (1)[a] | 6.50 | – | 1.36 | – | 7.11 | – | 1.12 | – |
| *sh2* (3) | 5.11 ± 0.58 | 3.87 ± 1.67 | 2.81 ± 0.75 | 1.52 ± 1.00 | 0.83 ± 0.14 | 0.60 ± 0.21 | 0.44 ± 0.12 | 0.72 ± 0.68 |
| *bt1* (1) | 5.67 | 4.83 | 4.50 | – | 6.08 | 1.92 | 1.58 | – |
| Plus RLG | | | | | | | | |
| *wx* (2) | 43.75 ± 6.1 | 37.78 ± 7.3 | 63.94 ± 0.7 | 58.49 ± 1.1 | – | – | – | – |
| *sh2* (2) | – | – | 47.56 ± 9.1 | 42.93 ± 9.1 | – | – | 2.49 ± 0.7 | 2.02 ± 0.6 |

[a] For this experiment the uptake solution contained 0.5 mM 3-PGA rather than the 12.4 mM used for all of the others.

**Table V.** *ADP-Glc uptake and incorporation into methanol- and water-insoluble products*

Amyloplasts were from the 20% Percoll fraction. The reaction mixture (200 μL final volume) contained 100 mM Bicine, pH 8.5, 0.5 M sorbitol, 12.5 mM EDTA, 10 mM GSH, 50 mM potassium acetate, and 4 mM [$^{14}$C]ADP-Glc plus 60 or 80 μL of amyloplast fraction. For the uptake studies amyloplasts were isolated from the endosperms of the following age kernels: N, two 10 DPP, one 11 DPP, and two 12 DPP; *wx*, one 10 DPP, three 12 DPP, five 13 DPP, and one 14 DPP; *sh2*, one each of 15 and 16 DPP; and *bt1*, one each of 11, 12, and 14 DPP. Data are averages ± SE.

| Genotype (*n*) | Total Incorporation | | Percentage Water Insoluble | |
|---|---|---|---|---|
| | Intact | Lysed | Intact | Lysed |
| | *pmol min$^{-1}$ 10$^{-6}$ amyloplasts* | | *% of total methanol insoluble* | |
| N (5) | 58.58 ± 24.5 | 12.33 ± 7.75 | 90 ± 6 | 87 ± 8 |
| *wx* (10) | 67.25 ± 16.83 | 6.17 ± 4.00 | 79 ± 6 | 80 ± 11 |
| *wx* (1)[a] | 49.37 | 6.29 | 94 | 95 |
| *sh2* (2) | 133.75 ± 8.42 | 42.92 ± 3.67 | 86 ± 10 | 92 ± 4 |
| *bt1* (3) | 18.00 ± 6.17 | 12.33 ± 4.17 | 67 ± 4 | 87 ± 1 |

[a] For this experiment the uptake solution contained 15 mM Hepes, pH 7.5, 0.08% BSA, 10 mM $MgCl_2$, 0.5 M sorbitol, and 4 mM ADP-Glc plus 80 μL of amyloplast suspension.

but incorporation by lysed *bt1* amyloplasts was similar to that by lysed normal and *wx* amyloplasts.

The addition of glycogen, an alternative glucosyl acceptor, to the reactions containing the lysed amyloplasts from *wx* and *bt1* genotypes restored incorporation equal to or greater than that in reactions containing intact *wx* amyloplasts (Table VI). Thus, the reduced [$^{14}$C]Glc incorporation from ADP-Glc into starch by *bt1* amyloplasts was caused by the reduced transfer of ADP-Glc into the amyloplasts. In addition, we can conclude from these results that the reduced incorporation from ADP-Glc by lysed amyloplasts in the absence of added glycogen acceptor was apparently the result of dilution of the SS relative to the nonreducing ends of the native maltooligosaccharide or starch-granule acceptors. Enhanced incorporation of [$^{14}$C]Glc from ADP-Glc in the presence of rabbit-liver glycogen (Table VI) indicates that glycogen may be a better substrate for SS than the native acceptors.

If an adenylate translocator in the amyloplast membrane of the isolated intact amyloplasts is functioning in the uptake of ADP-Glc in exchange for ADP or AMP, then these ADP or AMP nucleotides in the uptake medium might compete with ADP-Glc for uptake. In fact, preincubation of the amyloplasts in the cold for 30 min in the presence of 7 mM ADP or AMP, followed by [$^{14}$C]ADP-Glc uptake and incorporation in the presence of 2.2 mM ADP or AMP, reduced $^{14}$C incorporation into starch by 75% and 82%, respectively (data not shown). SS synthesis activities, as measured in reactions containing lysed amyloplasts plus glycogen, were inhibited 60% and 63%, respectively, by these ADP and AMP treatments. Therefore, in this study we were unable to distinguish between the effects of the nucleotides on ADP-Glc uptake and their effects on SS activity.

The adenosine analog FSBA is well known to react with adenosine nucleotide-binding sites of enzymes and proteins (Colman, 1983), including mitochondrial $F_1$-ATPase (Esch and Allison, 1978), chloroplast ATPase (DeBenedetti and Jagendorf, 1979), and an ADP-binding protein on the exterior surface of human platelets (for review, see Colman, 1983). If ADP-Glc is transported into amyloplasts via an adenylate translocator, then we predicted that FSBA would inhibit ADP-Glc uptake into intact amyloplasts, resulting in reduced incorporation of Glc into starch. To test this we pretreated intact amyloplasts from *wx* in the standard uptake medium containing 0 to 4 mM FSBA dissolved in DMSO. All pretreatment and uptake solutions contained 2% DMSO (the FSBA solvent), which had no negative effect on uptake of ADP-Glc and incorporation of [$^{14}$C]Glc into the methanol-insoluble product in either the absence or presence of rabbit-liver glycogen (incorporation in the presence of glycogen is a measure of SS activity) (Table VII).

Inhibition of uptake and incorporation by intact amyloplasts increased with increasing concentrations of FSBA. FSBA would also be expected to interact with the ADP-Glc-binding site of SS, but at 2 and 4 mM FSBA, reduction in uptake and incorporation into starch by intact amyloplasts was greater than the inhibition of SS as measured in lysed Table VII amyloplasts in the presence of added glycogen (Table VII). For intact amyloplasts incubated in the absence of glycogen (except at the highest FSBA treatment), almost three-fourths of the methanol-insoluble radioactivity was incorporated into the water-insoluble starch granules. The amyloplast-uptake studies provide evidence that cytosolic synthesized ADP-Glc can be transferred across amyloplast membranes via an adenylate translocator. We have proposed that BT1 is that adenylate translocator in maize endosperms (Shannon et al., 1996).

**Table VI.** *Effect of glycogen on ADP-Glc incorporation into methanol-insoluble products*

Amyloplast preparation and uptake conditions were the same as in Table V, with the addition of 1 mg of rabbit-liver glycogen (RLG) where noted. For each genotype, data are the means of duplicate incubations of a single preparation of amyloplasts from 13 DPP *wx* endosperm and from 14 DPP *bt1* endosperm.

| Genotype | Total Incorporation | | |
|---|---|---|---|
| | Intact | Lysed | Lysed + RLG |
| | *pmol min$^{-1}$ 10$^{-6}$ amyloplasts* | | |
| *wx* | 78.2 | 13.6 | 110.3 |
| *bt1* | 26.5 | 18.1 | 130.9 |

**Table VII.** *FSBA inhibition of ADP-Glc uptake and incorporation into methanol- and water-insoluble products and inhibition of SS*

Amyloplasts in the 20% Percoll fraction isolated from 13 DPP *wx* endosperms were used. The amyloplasts were preincubated for 30 min at 30°C in their respective reaction mixtures before the addition of [$^{14}$C]ADP-Glc. The reaction mixtures were as described in Table VI, with the addition of 2% DMSO and FSBA as noted. Reaction mixtures containing 1 mg of rabbit-liver glycogen (RLG) provide a measure of SS activity. Control intact and lysed amyloplasts, without DMSO, incorporated 34.62 and 4.63 pmol min$^{-1}$ 10$^{-6}$ amyloplasts, respectively.

| FSBA | RLG | Total Incorporation of [$^{14}$C]ADP-Glc | | Inhibition of Incorporation by FSBA | | Incorporated into Starch Granules | |
|---|---|---|---|---|---|---|---|
| | | Intact | Lysed | Intact | Lysed | Intact | Lysed |
| mM | | pmol min$^{-1}$ 10$^{-6}$ amyloplasts | | | | % | |
| 0 | − | 48.25 | 5.67 | 0 | 0 | 76 | 92 |
| 1 | − | | | 6 | 0 | 68 | 97 |
| 2 | − | | | 43 | 33 | 69 | 67 |
| 4 | − | | | 62 | 30 | 53 | 54 |
| 0 | + | 57.92 | 59.08 | 0 | 0 | 27 | 8 |
| 1 | + | | | 0 | 9 | 37 | 10 |
| 2 | + | | | 3 | 16 | 30 | 8 |
| 4 | + | | | 20 | 41 | 25 | 10 |

### *Identification of a Putative ADP-Glc-Binding Motif in BT1*

If BT1 is the adenylate translocator functioning in the transfer of ADP-Glc into amyloplasts, then BT1 must contain an ADP-Glc-binding motif. Analysis of the full-length BT1 sequence showed the presence of a KTGGL motif. This motif was identified as the ADP-Glc-binding site of *Escherichia coli* glycogen synthase (Furukawa et al., 1993) and is conserved in all known enzymes that use ADP-Glc as a substrate, including plant SS and bacterial glycogen synthases (Table VIII). The KTGGL motif in BT1 is 40 amino acid residues upstream of the transit-peptide cleavage site proposed by Sullivan et al. (1991). Thus, if this proposed ADP-Glc-binding motif is present in the mature BT1 protein, then an alternative transit-peptide cleavage site is required. Comparison of several known N-terminal sequences of SS revealed a consensus cleavage site of V(I)X/A(G,S), and in BT1 an alternative cleavage site, VP/A, is present 13 amino acid residues upstream of the KTGGL motif, the proposed ADP-Glc-binding site (Table IX). Cleavage at this site would yield a mature BT1 protein of 44

**Table VIII.** *Positions of ADP-Glc-binding motifs in the full-length BT1 protein and in various starch (SS) and glycogen synthases (GS)*

| Protein | Accession No. | Motif | Position | Reference |
|---|---|---|---|---|
| Maize BT1 | M79333 | KTGGL<br>V<br>A | 35–39 | Sullivan et al. (1991) |
| ADP-Glc-binding consensus sequence | | S<br>KTGGL | | |
| Barley WX | X07932 | KTGGL | 90–94 | Rohde et al. (1988) |
| Cassava WX | X74160 | KTGGL | 96–100 | Salehuzzaman et al. (1993) |
| Maize WX | M24258 | KTGGL | 90–94 | Kloesgen et al. (1986) |
| Pea GBSSI | X88789 | KTGGL | 91–95 | Dry et al. (1992) |
| Pea GBSSII | X88790 | KTGGL | 255–259 | Dry et al. (1992) |
| Potato SS | X87988 | KTGGL | 360–364 | Edwards et al. (1995) |
| Potato SSIII | X94440 | KVGGL | 794–798 | Abel et al. (1996) |
| Potato SSSI | Y10416 | KTGGL | 145–149 | (G.I.W. Abel, J. Kossman, and L. Willmitzer, unpublished data) |
| Potato SSSIII | X95759 | KVGGL | 794–798 | Marshall et al. (1996) |
| Potato WX | X58453 | KTGGL | 95–99 | van der Leij et al. (1991) |
| Rice SSS | D16202 | KSGGL | 97–101 | Baba et al. (1993) |
| Rice WX | X62134 | KTGGL | 97–101 | Okagaki (1992) |
| Sorghum WX | U23945 | KTGGL | 96–100 | Hsieh et al. (1996) |
| Sweet potato SS | U44126 | KTGGL | 98–102 | (S.-J. Wang, K.W. Yeh, and C.-Y. Tsai, unpublished data) |
| Wheat WX | X57233 | KTGGL | 91–95 | Clark et al. (1991) |
| *E. coli* GS | J02616 | KTGGL | 15–19 | Kumar et al. (1986) |
| *Bacillus stearothermophilus* GS | D87026 | KSGGL | 15–19 | (H. Takata, T. Takata, S. Okada, M. Takagi, and T. Imanaka, unpublished data) |
| *Bacillus subtilis* GS | Z25795 | KSGGL | 15–19 | Kiel et al. (1994) |
| *Agrobacterium tumefaciens* GS | L24117 | KTGGL | 15–19 | Uttaro and Ugalde (1994) |
| *Synechocystis* GS | D90899 | KAGGL | 15–19 | Kaneko et al. (1996) |

**Table IX.** *Positions of putative transit peptide cleavage sites in the full-length BT1 protein and in various starch synthases (SS)*

| Protein | Accession No. | Cleavage Site | Method | Position | Ref. |
|---|---|---|---|---|---|
| Maize BT1 | M79333 | SLQVP/AV·<br>S | Sequence alignment | 24/25 | Sullivan et al. (1991) |
| Transit peptide cleavage site | | I G | | | |
| Consensus sequence | | VX/A | | | |
| Barley WX | X07932 | SVVVS/AT | Sequence alignment | 70/71 | Rohde et al. (1988) |
| Cassava WX | X74160 | AKIIC/GH | Sequence alignment | 78/79 | Salehuzzaman et al. (1993) |
| Maize WX | M24258 | SLVVC/AS | Protein sequencing | 72/73 | Kloesgen et al. (1986) |
| Pea GBSSI | X88789 | GKIVC/GM | Protein sequencing | 75/76 | Dry et al. (1992) |
| Pea GBSSII | X88790 | KQHVR/AV | Protein sequencing | 57/58 | Dry et al. (1992) |
| Potato SS | X87988 | NQRVK/AT | Protein sequencing | 65/66 | Edwards et al. (1995) |
| Potato WX | X58453 | ATIVC/GK | Protein sequencing | 77/78 | van der Leij et al. (1991) |
| Rice SSS | D16202 | TIFVA/SE | Protein sequencing | 24/25 | Baba et al. (1993) |
| Rice WX | X62134 | SVVVY/AT | Protein sequencing | 77/78 | Okagaki (1992) |
| Sorghum WX | U23945 | SLVVC/AT | Sequence alignment | 77/78 | Hsieh et al. (1996) |
| Wheat WX | X57233 | SMVVR/AT | Protein sequencing | 70/71 | Clark et al. (1991) |

kD, which agrees well with the 39 to 44 kD for BT1 reported previously (Cao et al., 1995).

## DISCUSSION

In this paper we summarize the results of several studies that strongly support the conclusion that in maize endosperm most of the cellular AGPase is localized in the cytosol, and that the inner amyloplast-membrane-specific polypeptide, BT1, is an adenylate translocator that functions in the transfer of cytosol-synthesized ADP-Glc into the amyloplasts. Denyer et al. (1996) reported that more than 95% of AGPase activity in maize endosperm cells is extraplastidial. This result was based on aqueous fractionation of endosperm homogenates from young (11–17 DPP) kernels. During fractionation approximately 75% of the amyloplast marker enzymes were lost from the amyloplast fraction (Denyer et al., 1996), and we know from experience that a greater proportion of the more mature amyloplasts with the larger starch granules are lysed during isolation. Consequently, the final preparation would be enriched with amyloplasts containing small starch granules. Enzyme compartmentation in such an amyloplast preparation may not be representative of compartmentation in the more mature amyloplasts from cells in the linear phase of starch accumulation.

This concern is validated by a recent immunolocalization study by Brangeon et al. (1997), which clearly shows that the peripheral endosperm cells were only lightly immunolabeled by antibodies to BT2 and SH2 and that there was a gradient of increasing signal intensity that paralleled the increase in starch-granule size. We have developed and used nonaqueous TCE/heptane fractionation and nonaqueous glycerol-isolation methods to show for the first time, to our knowledge, that in the more mature maize endosperm cells (20 DPP), 90% or more of the cellular AGPase is cytosolic (Tables I and II). These studies were based on the observation that during freeze-drying, much of the amyloplast stromal content dries onto the surface of the starch granule, and the starch granule also shrinks away from the cytosol (Liu and Shannon, 1981). Therefore, during nonaqueous fractionation (Table I) or isolation (Table II) the stromal enzymes associated with the starch granules remain with the granules until they are extracted with the aqueous buffer solution.

With the TCE/heptane protocol, patterned after the procedures used by Riens et al. (1991) and MacDougall et al. (1995), pulverized endosperm samples were separated into several fractions with varying enrichments in amyloplasts or cytosol. With this procedure enzyme compartmentation in amyloplasts was determined from plots of the activity of the enzyme in question per million starch granules against the activity of nonplastidial enzymes per million granules. This procedure has the advantage that enzyme recovery was high after TCE/heptane fractionation. It was clear from these data that AGPase closely partitioned with the nonplastidial marker enzymes (Fig. 1), and 58% to 77% of the plastid marker-enzyme activities were retained with the starch granules (Table I).

Although the TCE/heptane procedure could be used to estimate enzyme compartmentation in amyloplasts, it was not satisfactory for the isolation of a cytosol-free starch-granule preparation with associated stromal enzymes (amyloplasts). To accomplish this we modified the glycerol-based procedure of Liu and Shannon (1981). The resulting starch-granule preparations, which were contaminated with about 7% of the cytosol marker enzymes, retained approximately 14% of the cellular AGPase (7% more than cytosolic contamination) and 50% of the amyloplast marker-enzyme activities (Table II). It is significant that enzyme activities per million starch granules determined by both nonaqueous procedures were approximately the same (Tables I and II). Results of the nonaqueous studies that demonstrated predominant cytosolic localization of AGPase were confirmed by immunolocalization of BT2 and SH2 polypeptides in aqueously isolated amyloplasts (Fig. 4).

Earlier studies of compartmentation of AGPase in maize endosperm cells have been controversial. In contrast to the extra-amyloplastic localization of AGPase reported by

Denyer et al. (1996), results of immunocytolocalization studies have been interpreted as showing that most, if not all, AGPase is localized in the amyloplasts (Miller and Chourey, 1995; Brangeon et al., 1997). Both approaches to the study of compartmentation in maize endosperm have serious disadvantages. First, as noted above, 75% or more of the amyloplasts are ruptured during aqueous isolation, resulting in an amyloplast preparation enriched in plastids containing the smaller starch granules (Shannon, 1987). Second, a serious drawback of immunocytolocalization studies of maize endosperm at the electron-microscopic level is the difficulty of sufficiently embedding the tissue so that the thin slices of starch granules remain in the plastic. Consequently, regions of the endosperm consisting of cells with small starch granules are more likely to survive preparation. In maize endosperms such cells occur in very young kernels (about 12 DPP) or in the peripheral cells.

As noted above, enzyme compartmentation in these physiologically less mature cells may not be representative of compartmentation in cells more actively engaged in starch synthesis (Tsai et al., 1970; Brangeon et al., 1997). In addition, it is possible that cytosolic enzymes may be lost from the tissue piece during preparation of the samples for electron-microscopic examination; in fact, Miller and Chourey (1995) pointed out that they were unable to immunolocalize the cytosol-specific enzyme SUS. Tissue preparation and cutting of the thicker sections suitable for immunocytolocalization at the light-microscopic level are less problematic than preparation for electron-microscopic studies. In a light-microscopic immunolocalization study of AGPase compartmentation in maize kernels, Brangeon et al. (1997) clearly showed that in pericarp cells of kernels 8 DPP, polypeptides recognized by antibodies to the AGPase subunits BT2 and SH2 were cytosolic, but in endosperm cells actively engaged in starch synthesis (15 DPP), the antibodies immunolabeled only polypeptides that were closely associated with amyloplasts. These authors suggested an intraplastidial localization for the AGPase polypeptides encoded by *Bt2* and *Sh2* in these maize endosperm cells. However, they suggest that at this level of resolution, it is not possible to distinguish between proteins in the amyloplast stroma and proteins located either between the inner and outer membranes of the amyloplast envelope or closely associated (loosely bound) with the outer membrane (Brangeon et al., 1997).

The compartmentation results obtained in the present study using nonaqueous procedures do not support the suggestion of Brangeon et al. (1997) that AGPase is localized in the amyloplast stroma. Rather, we show that AGPase resides in a compartment that partitions with the cytosol during nonaqueous fractionation. However, because the soluble enzymes located within the inner-membrane space of the amyloplast envelopes and those in close association with the amyloplasts in situ would be expected to partition with the cytosol during aqueous and nonaqueous fractionation or isolation, we were unable to rule out the possibility that AGPase resides within the inner-membrane space of amyloplasts.

### BT1 Is an Adenylate Translocator

Two phosphate translocators (Fischer et al., 1997; Kammerer et al., 1998) and two adenylate translocators (Möhlmann et al., 1997) have been reported to be present in maize endosperm amyloplast membranes. Fischer et al. (1997) isolated and characterized a PEP/Pi antiporter that is present in plastid membranes from both photosynthetic and nonphotosynthetic tissues. The Glc-6-P/Pi antiporter was shown to be preferentially expressed in nonphotosynthetic tissues (Kammerer et al., 1998) and to mediate the 1:1 exchange of Glc-6-P with Pi and triose phosphate, and is assumed to function in vivo in the import of Glc-6-P into amyloplasts. Kammerer et al. (1998) suggest that Glc-6-P may be used either in the starch biosynthetic pathway or as a substrate for the oxidative pentose-phosphate pathway. Neuhaus et al. (1993) reported the isolation of amyloplasts from maize endosperm that were capable of uptake and incorporation Glc-6-P into starch. Möhlmann et al. (1997) used a similar amyloplast-isolation procedure and determined that Glc from ADP-Glc was incorporated into starch at a rate 6 times higher than that from Glc-6-P.

The amyloplast-isolation procedure used for these studies, which included multiple high-speed centrifugations through density gradients, yielded a preparation of amyloplasts with very small starch granules (Neuhaus et al., 1993) and most likely also contained amyloplast membrane vesicles without starch granules. We have used a much more gentle amyloplast-isolation procedure, and the results of the hexose-P uptake and incorporation studies presented in this paper do not support the use of Glc-6-P in the starch biosynthetic pathway (Table IV). Intact amyloplasts from maize endosperm were relatively inefficient in the uptake and conversion of Glc-1-P and Glc-6-P into starch regardless of whether ATP was included in the uptake medium. In this study we measured incorporation of radioactive hexoses into starch and did not attempt to determine hexose-P uptake independent of its utilization in starch synthesis. If we assume that the amyloplast membranes contain a functional Glc-6-P/Pi antiporter, then the imported Glc-6-P is a poor substrate for starch synthesis. The amyloplasts used for these uptake and incorporation studies were isolated from young kernels that may not have developed their full complement of AGPase activity (Tsai et al., 1970; Brangeon et al., 1997), and this may have contributed to the poor utilization of hexose-Ps. However, when considering the predominant cytosolic localization of AGPase (Table I and II), the poor hexose-P utilization may simply reflect the minor role of amyloplastic AGPase in starch synthesis.

Amyloplasts isolated from *wx* and *sh2* endosperms apparently do contain active starch phosphorylase, because when glycogen, an alternative glucan acceptor, was included in the uptake mixture Glc was effectively transferred from Glc-1-P to the glycogen acceptor (Table IV). In contrast, Glc-6-P was a poor substrate for Glc addition to glycogen, indicating either that the amyloplasts from *sh2* contain very little phosphoglucomutase or that it is inactive in the incubation conditions used.

Results of in vivo studies of the starch-deficient maize endosperm mutant *bt1* support the conclusion that BT1, an amyloplast-membrane-specific polypeptide (Cao et al., 1995; Sullivan and Kaneko, 1995), is an adenylate translocator that functions in ADP-Glc transfer into amyloplasts. For example, ADP-Glc, which is synthesized by AGPase, accumulates in the endosperm of *bt1* mutant kernels (Shannon et al., 1996). Activities of AGPase, UGPase, SS, extractable SBE, and SUS in extracts from *bt1* mutant endosperms were equal to or greater than activities in endosperm extracts from normal kernels (Shannon et al., 1996). The genetic lesion in *bt1* kernels was found to be an amyloplast-membrane-specific, 39- to 44-kD polypeptide, BT1 (Cao et al., 1995; Sullivan and Kaneko, 1995).

Based on these results, we suggest that BT1 is an adenylate translocator that functions in the transfer of ADP-Glc from the cytosol into the amyloplast, and in its absence (i.e. in *bt1* mutant kernels) ADP-Glc accumulates (Shannon et al., 1996). The most direct support for this suggestion is provided by the marked difference in the uptake of ADP-Glc and its use for starch synthesis by amyloplasts isolated from *bt1* endosperms and amyloplasts isolated from normal, *wx*, and *sh2* endosperms (Table V). Intact amyloplasts from *bt1* endosperms, which are missing the BT1 polypeptides, were only 26% as effective in taking up and converting ADP-Glc to starch as those from the other genotypes (Table V).

Several lines of evidence support the conclusion that we were measuring ADP-Glc uptake and utilization by intact amyloplasts and not simply synthesis by SS associated with granules released from lysed amyloplasts: (a) ADP-Glc incorporation by lysed amyloplasts was only about 10% of that by intact amyloplasts; (b) for many of the uptake and incorporation studies we used amyloplasts from the *wx* endosperm mutant, which is deficient in the starch-granule-bound starch synthase (Shannon and Garwood, 1984); and (c) the adenosine analog FSBA, which is known to react with adenosine-binding sites (Colman, 1983), more effectively inhibited uptake and incorporation of ADP-Glc than starch synthase (Table VIII).

Comparison of the translated full-length sequence of BT1 with protein sequences of 45 adenylate translocators from 20 species revealed about 30% identity and 81% similarity within the highly conserved regions of the mitochondrial adenylate translocators (Cao and Cao, 1997). If BT1 is an adenylate translocator, as was suggested, then the mature protein should contain an ADP-Glc-binding motif. However, no ADP-Glc-binding motif was present in the mature BT1 protein, assuming that the transit-peptide cleavage site VRA/A that was proposed by Sullivan et al. (1991) is correct. However, analysis of the full-length BT1 amino acid sequence showed the presence of the putative ADP-Glc-binding motif, KTGGL, 40 amino acid residues upstream of the cleavage site proposed by Sullivan et al. (1991) (Table VIII). For BT1 we propose an alternative transit-peptide cleavage site, VP/A, 13 amino acids upstream of the putative ADP-Glc-binding motif (Table IX). Transit-peptide cleavage at this site would yield a mature BT1 of 44 kD, which agrees well with the size we reported for mature BT1 (Cao et al., 1995) but is somewhat larger than the 39.5- and 38.5-kD mature BT1 polypeptides reported by Li et al. (1992).

In summary, we have provided evidence that most of the cellular AGPase in maize kernels in both the linear phase and in the early phase of starch accumulation resides in a compartment that partitions with cytosolic marker enzymes after nonaqueous and aqueous fractionation. However, based on the immunolocalization study of Brangeon et al. (1997), we suggest that in situ AGPase is functionally compartmented with the amyloplasts and may be loosely associated with the outer membrane of the amyloplast envelope. ADP-Glc is transported into the amyloplast stroma via BT1, which may be the same transporter as the ADP-Glc/AMP adenylate translocator described by Möhlmann et al. (1997). The importance of the BT1 translocator to starch accumulation in maize endosperms is demonstrated by the severely reduced starch content in *bt1* mutant kernels (Tobias et al., 1992). Assessment of the relative importance of the hexose-P/Pi antiporter for starch accumulation in vivo awaits isolation of a mutant genotype defective in the hexose-P/Pi antiporter.

## ACKNOWLEDGMENTS

The authors acknowledge generous gifts of polyclonal antibodies to maize SH2 and BT2 from Dr. L. Curtis Hannah (Department of Horticultural Sciences, University of Florida, Gainesville) and antibodies to maize endosperm BT1 from Dr. Thomas D. Sullivan (Laboratory of Genetics, University of Wisconsin-Madison).

Received February 27, 1998; accepted April 24, 1998.


## LITERATURE CITED

**Abel GJ, Springer F, Willmitzer L, Kossmann J** (1996) Cloning and functional analysis of a cDNA encoding a novel 139 kDa starch synthase from potato (*Solanum tuberosum* L.). Plant J **10:** 981–991

**Baba T, Nishihara M, Mizuno K, Kawasaki T, Shimada H, Kobayashi E, Ohnishi S, Tanaka K, Arai Y** (1993) Identification, cDNA cloning, and gene expression of soluble starch synthase in rice (*Oryza sativa* L.) immature seeds. Plant Physiol **103:** 565–573

**Boller T, Kende H** (1979) Hydrolytic enzymes in the central vacuole of plant cells. Plant Physiol **63:** 1123–1132

**Boyer CD, Shannon JC, Hardison RC, eds** (1989) Physiology, Biochemistry, and Genetics of Nongreen Plastids. Current Topics in Plant Physiology: An American Society of Plant Physiologists Series, Vol 2. American Society of Plant Physiologists, Rockville, MD

**Brangeon J, Reyss A, Prioul J-L** (1997) *In situ* detection of ADP-glucose pyrophosphorylase expression during maize endosperm development. Plant Physiol Biochem **35:** 847–858

**Briskin DP, Leonard RT, Hodges TK** (1987) Isolation of the plasma membrane: membrane markers and general principles. Methods Enzymol **148:** 542–558

**Cao H, Cao J** (1997) Comparison of forty-five adenylate translocator sequences from twenty different species supports the hypothesis that Brittle-1 protein is an adenylate translocator in maize amyloplast membranes (abstract no. 895). Plant Physiol **114:** S-181

**Cao H, Preiss J** (1996) Evidence for essential arginine residues at the active sites of maize branching enzymes. J Protein Chem **15:** 291–304

**Cao H, Shannon JC** (1996) BT1, a protein critical for *in vivo* starch accumulation in maize endosperm, is not detected in maize endosperm suspension cultures. Physiol Plant **97:** 665–673

**Cao H, Shannon JC** (1997) BT1, a possible adenylate translocator, is developmentally expressed in maize endosperm but not detected in starchy tissues from several other species. Physiol Plant **100:** 400–406

**Cao H, Sullivan TD, Boyer CD, Shannon JC** (1995) *Bt1*, a structural gene for the major 39–44 kDa amyloplast membrane polypeptides. Physiol Plant **95:** 176–186

**Clark JR, Robertson M, Ainsworth CC** (1991) Nucleotide sequence of a wheat (*Triticum aestivum* L.) cDNA clone encoding the waxy protein. Plant Mol Biol **16:** 1099–1101

**Colman RF** (1983) Affinity labeling of purine nucleotide sites in proteins. Annu Rev Biochem **52:** 67–91

**DeBenedetti E, Jagendorf A** (1979) Inhibition of chloroplast coupling factor ATPase by 5′-*p*-fluorosulfonylbenzoyl adenosine. Biochem Biophys Res Commun **86:** 440–446

**Denyer K, Dunlap F, Thorbjornsen T, Keeling P, Smith AM** (1996) The major form of ADP-glucose pyrophosphorylase in maize endosperm is extra-plastidial. Plant Physiol **112:** 779–785

**Dry I, Smith A, Edwards A, Bhattacharyya M, Dunn P, Martin C** (1992) Characterization of cDNAs encoding two isoforms of granule-bound starch synthase which show differential expression in developing storage organs of pea and potato. Plant J **2:** 193–202

**Echeverria E, Boyer CD, Liu K-C, Shannon JC** (1985) Isolation of amyloplasts from developing maize endosperm. Plant Physiol **77:** 513–519

**Echeverria E, Boyer CD, Thomas PA, Liu K-C, Shannon JC** (1988) Enzyme activities associated with maize kernel amyloplasts. Plant Physiol **86:** 786–792

**Edwards A, Marshall J, Sidebottom C, Visser RG, Smith AM, Martin C** (1995) Biochemical and molecular characterization of a novel starch synthase from potato tubers. Plant J **8:** 283–294

**Esch FS, Allison WS** (1978) Identification of a tyrosine residue at a nucleotide binding site in the β-subunit of the mitochondrial ATPase with *p*-fluorosulfonyl[$^{14}$C]-benzoyl-5′-adenosine. J Biol Chem **253:** 6100–6106

**Fischer K, Kammerer B, Gutensohn M, Arbinger B, Weber A, Häusler RE, Flügge U-I** (1997) A new class of plastidic phosphate translocators: a putative link between primary and secondary metabolism by the phosphoenolpyruvate/phosphate antiporter. Plant Cell **9:** 453–462

**Fisher DK, Gao M, Kim K-N, Boyer CD, Guiltinan MJ** (1996) Allelic analysis of the maize *amylose-extender* locus suggests that independent genes encode starch-branching enzymes IIa and IIb. Plant Physiol **110:** 611–619

**Furukawa K, Tagaya M, Tanizawa K, Fukui T** (1993) Role of the conserved lys-X-gly-gly sequence at the ADP-glucose-binding site in *Escherichia coli* glycogen synthase. J Biol Chem **268:** 23837–23842

**Gardner HW, Miernyk JA, Christianson DD, Khoo U** (1987) Isolation and characterization of an amyloplast envelope-enriched fraction from immature maize endosperm. Physiol Plant **69:** 541–549

**Giroux MJ, Hannah LC** (1994) ADP-glucose pyrophosphorylase in *shrunken-2* and *brittle-2* mutants of maize. Mol Gen Genet **243:** 400–408

**Gross P, ap Rees T** (1986) Alkaline inorganic pyrophosphatase and starch synthesis in amyloplasts. Planta **167:** 140–145

**Hannah LC, Giroux M, Boyer C** (1993) Biotechnological modification of carbohydrates for sweet corn and maize improvement. Sci Hortic **55:** 177–197

**Hsieh J, Liu C, Hsing YC** (1996) Molecular cloning of a sorghum cDNA encoding the seed waxy protein (PGR 96–119). Plant Physiol **112:** 1735

**James MG, Robertson DS, Myers AM** (1995) Characterization of the maize gene *sugary1*, a determinant of starch composition in kernels. Plant Cell **7:** 417–429

**Kammerer B, Fischer K, Hilpert B, Schubert S, Gutensohn M, Weber A, Flügge U-I** (1998) Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose 6-phosphate/phosphate antiporter. Plant Cell **10:** 105–117

**Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, and others** (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res **3:** 109–136

**Kiel JA, Boels JM, Beldman G, Venema G** (1994) Glycogen in *Bacillus subtilis*: molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation. Mol Microbiol **11:** 203–218

**Kloesgen RB, Gierl A, Schwarz-Sommer ZS, Saedler H** (1986) Molecular analysis of the *waxy* locus of *Zea mays*. Mol Gen Genet **203:** 237–244

**Kumar A, Larsen CE, Preiss J** (1986) Biosynthesis of bacterial glycogen: primary structure of *Escherichia coli* ADP-glucose:α-1,4-glucan, 4-glucosyltransferase as deduced from the nucleotide sequence of the *glgA* gene. J Biol Chem **261:** 16256–16259

**Li H-M, Sullivan TD Keegstra K** (1992) Information for targeting to the chloroplastic inner envelope membrane is contained in the mature region of the maize *Bt1*-encoded protein. J Biol Chem **267:** 18999–19004

**Liu K-C, Boyer CD, Shannon JC** (1992) Carbohydrate transfer into isolated maize (*Zea mays* L.) amyloplasts (abstract no. 234). Plant Physiol **99:** S-39

**Liu T-TY, Shannon, JC** (1981) A nonaqueous procedure for isolating starch granules with associated metabolites from maize (*Zea mays* L.) endosperm. Plant Physiol **67:** 518–524

**MacDougall AJ, Parker R, Selvendran RR** (1995) Nonaqueous fractionation to assess the ionic composition of the apoplast during fruit ripening. Plant Physiol **108:** 1679–1689

**Marshall J, Sidebottom C, Debet M, Martin C, Smith AM, Edwards A** (1996) Identification of the major starch synthase in the soluble fraction of potato tubers. Plant Cell **8:** 1121–1135

**Miller ME, Chourey PS** (1995) Intracellular immunolocalization of adenosine 5′-diphosphate-glucose pyrophosphorylase in developing endosperm cells of maize (*Zea mays* L.). Planta **197:** 522–527

**Möhlmann T, Tjaden J, Henrichs G, Quick WP, Häusler R, Neuhaus HE** (1997) ADP-glucose drives starch synthesis in isolated maize endosperm amyloplasts: characterization of starch synthesis and transport properties across the amyloplast envelope. Biochem J **324:** 503–509

**Mu-Forster C, Huang R, Powers JR, Harriman RW, Knight M, Singletary GW, Keeling PL, Wasserman BP** (1996) Physical association of starch biosynthetic enzymes with starch granules of maize endosperm. Granule-associated forms of starch synthase I and starch branching enzyme II. Plant Physiol **111:** 821–829

**Neuhaus HE, Batz O, Thom E, Scheibe R** (1993) Purification of highly intact plastids from various heterotrophic plant tissues: analysis of enzymic equipment and precursor dependency for starch biosynthesis. Biochem J **296:** 395–401

**Oh-Lee T-M, Setter TL** (1985) Enzyme activities of starch and sucrose pathways and growth of apical and basal maize kernels. Plant Physiol **79:** 848–851

**Okagaki RJ** (1992) Nucleotide sequence of a long cDNA from the rice gene. Plant Mol Biol **19:** 513–516

**Okita TW** (1992) Is there an alternative pathway for starch synthesis? Plant Physiol **100:** 560–564

**Oliveira L, Burns A, Bisalputra T, Yang K-C** (1983) The use of an ultra-low viscosity medium (VCD/HXSA) in the rapid embedding of plant cells for electron microscopy. J Microsc **132:** 195–202

**Pien F-M, Shannon JC** (1996) Cytosol versus plastid localization of starch biosynthetic enzymes in developing maize endosperm: evidence for the localization of ADP-Glc pyrophosphorylase in the cytosol (abstract no. 382). Plant Physiol **102:** S-100

**Pozueta-Romero J, Frehner M, Viale AM, Akazawa T** (1991) Direct transport of ADP glucose by an adenylate translocator is linked to starch biosynthesis in amyloplast. Proc Natl Acad Sci USA **88:** 5769–5773

**Preiss J** (1991) Biology and molecular biology of starch synthesis and its regulation. Oxf Surv Plant Mol Cell Biol **7:** 59–114

**Rahman A, Wong K-S, Jane J, Myers AM, James MG** (1998) Characterization of SU1 isoamylase, a determinant of storage starch structure in maize. Plant Physiol **117:** 425–435

**Riens B, Lohaus G, Heineke D, Heldt HW** (1991) Amino acid and sucrose content determined in the cytosolic, chloroplastic, and vacuolar compartments and in the phloem sap of spinach leaves. Plant Physiol **97:** 227–233

**Rohde W, Becker D, Salamini F** (1988) Structural analysis of the *waxy* locus from *Hordeum vulgare*. Nucleic Acids Res **16:** 7185–7186

**Salehuzzaman SN, Jacobsen E, Visser RG** (1993) Isolation and characterization of a cDNA encoding granule-bound starch synthase in cassava (*Manihot esculenta* Crantz) and its antisense expression in potato. Plant Mol Biol **23:** 947–962

**Shannon JC** (1989) Aqueous and nonaqueous methods for amyloplast isolation. *In* CD Boyer, JC Shannon, RC Hardison, eds, Physiology, Biochemistry, and Genetics of Nongreen Plastids. Current Topics in Plant Physiology: An American Society of Plant Physiologists Series, Vol 2. American Society of Plant Physiologists, Rockville, MD, pp 37–48

**Shannon JC, Echeverria E, Boyer CD** (1987) Isolation of amyloplasts from developing endosperm of maize (*Zea mays* L.). Methods Enzymol **148:** 226–234

**Shannon JC, Garwood DL** (1984) Genetics and physiology of starch development. *In* RL Whistler, JN Bemiller, EP Pascall, eds, Starch: Chemistry and Technology. Academic Press, New York, pp 25–86

**Shannon JC, Pien F-M, Liu K-C** (1996) Nucleotides and nucleotide sugars in developing maize (*Zea mays* L.) endosperms. Synthesis of ADP-glucose in *brittle-1*. Plant Physiol **110:** 835–843

**Smith AM, Denyer K, Martin CR** (1997) The synthesis of the starch granule. Annu Rev Plant Physiol Plant Mol Biol **48:** 67–87

**Sullivan TD, Kaneko Y** (1995) The maize *brittle1* gene encodes amyloplast membrane polypeptides. Planta **196:** 477–484

**Sullivan TD, Strelow LI, Illingworth CA, Phillips RL, Nelson OE Jr** (1991) Analysis of maize *Brittle-1* alleles and a defective suppressor-mutator-induced mutable allele. Plant Cell **3:** 1337–1348

**Thorbjornsen T, Villand P, Denyer K, Olsen O-A, Smith AM** (1996) Distinct isoforms of ADPglucose pyrophosphorylase occur inside and outside the amyloplasts in barley endosperm. Plant J **10:** 243–250

**Tobias RB, Boyer CD, Shannon JC** (1992) Alteration in carbohydrate intermediates in the endosperm of starch-deficient maize (*Zea mays* L.) genotypes. Plant Physiol **99:** 146–152

**Tsai CY, Salamini F, Nelson OE** (1970) Enzymes of carbohydrate metabolism in developing endosperm of maize. Plant Physiol **46:** 299–306

**Uttaro AD, Ugalde RA** (1994) A chromosomal cluster of genes encoding ADP-glucose synthetase, glycogen synthase and phosphoglucomutase in *Agrobacterium tumefaciens*. Gene **150:** 117–122

**van der Leij FR, Visser RG, Ponstein AS, Jacobsen E, Feenstra WJ** (1991) Sequence of the structural gene for granule-bound starch synthase of potato (*Solanum tuberosum* L.) and evidence for a single point deletion in the *amf* allele. Mol Gen Genet **223:** 240–248

**Villand P, Kleczkowski LA** (1994) Is there an alternative pathway for starch biosynthesis in cereal seeds? Z Naturforsch **49c:** 215–219

Protein Engineering vol.9 no.1 pp.77–83, 1994

# Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide

**L.You[1] and F.H.Arnold[2]**

Division of Chemistry and Chemical Engineering 210–41, California Institute of Technology, Pasadena, CA 91125, USA

[1]Present address: Bristol-Myers Squibb Co., Bio/Chem Division, PO Box 4744, Syracuse, NY 13221, USA

[2]To whom correspondence should be addressed

**Sequential rounds of error-prone PCR to introduce random mutations and screening of the resultant mutant libraries have been used to enhance the total catalytic activity of subtilisin E significantly in a non-natural environment, aqueous dimethylformamide (DMF). Seven DNA substitutions coding for three new amino acid substitutions were identified in a mutant isolated after two additional generations of directed evolution carried out on 10M subtilisin E, previously 'evolved' to increase its specific activity in DMF. A *Bacillus subtilis–Escherichia coli* shuttle vector was developed in order to increase the size of the mutant library that could be established in *B.subtilis* and the stringency of the screening process was increased to reflect total as well as specific activity. This directed evolution approach has been extremely effective for improving enzyme activity in a non-natural environment: the resulting-evolved 13M subtilisin exhibits specific catalytic efficiency towards the hydrolysis of a peptide substrate succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide in 60% DMF solution that is three times that of the parent 10M and 471 times that of wild type subtilisin E. The total activity of the 13M culture supernatant is enhanced 16-fold over that of the parent 10M.**

*Keywords*: *B.subtilis*/directed evolution/subtilisin E/total activity/random mutagenesis/organic solvents

## Introduction

Available natural enzyme resources are being tailored to fulfil increasing demands for new biocatalysts. A practical strategy for altering enzyme properties is to introduce random base substitutions into the gene sequence and then select or screen for variants that express the desired phenotype(s). Features that have been enhanced by random mutagenesis include catalytic activity (Graham *et al.*, 1993), activity in organic solvents (Chen and Arnold, 1993), thermostability (Bryan *et al.*, 1986; Liao *et al.*, 1986; Joyet *et al.*, 1992), alkaline stability (Cunningham and Wells, 1987) and substrate specificity (Oliphant and Struhl, 1989; Graham *et al.*, 1993). Once the genes are sequenced, effective mutations can be accumulated by site-directed mutagenesis (Joyet *et al.*, 1992; Strausberg *et al.*, 1995). An attractive alternative to sequencing and site-directed mutagenesis is to accumulate beneficial mutations in sequential rounds of random mutagenesis (Chen and Arnold, 1993) or by a novel recombination approach (Stemmer, 1994), following a 'directed evolution' strategy (Arnold, in press). Directed evolution is likely to prove useful in enhancing enzyme performance in 'non-natural' environments (Arnold, 1993) as well as for obtaining new features never required by nature, provided an efficient selection or screening method can be found to channel the enzyme's evolution towards the desired properties. A significant advantage of this approach over 'rational' design methods is that neither structural information nor a mechanistic road-map are required to guide the directed evolution experiment.

In addition to creating new enzymes for applications in biotechnology, directed evolution methods can be used to explore the limits of protein function. Although naturally occurring enzymes have evolved within the context of living organisms, there is no reason to believe that they are limited to the relatively narrow set of conditions that will support life. Water-miscible organic solvents such as dimethylformamide (DMF), for example, are highly toxic to most organisms, even at levels of <10%. Although enzymes can often tolerate much higher concentrations and still retain their folded structures, catalytic activity is often significantly compromised. To probe whether enzymes could be 'tuned' to tolerate and even thrive in an apparently hostile environment (high concentrations of DMF), we applied a directed evolution strategy involving sequential generations of polymerase chain reaction (PCR) random mutagenesis and screening to the serine protease subtilisin E (Chen and Arnold, 1993). Mutations that improved the specific activity of subtilisin in aqueous DMF were not rare; mutations conferring ~2-fold enhancements in specific proteolytic activity could be found by screening only a few hundred colonies from each generation (Chen and Arnold, 1993). The accumulation of effective mutations in sequential rounds of random mutagenesis led to a significant enhancement of the enzyme's activity: the specific activity of the 'evolved' enzyme containing 10 amino acid substitutions (10M subtilisin E) was enhanced 157-fold in 60% DMF, recovering much of the activity lost by addition of the organic solvent.

Subtilisin is a useful catalyst for organic synthesis, particularly in the presence of organic solvents. Subtilisin can catalyze regioselective (Riva and Klibanov, 1988; Riva *et al.*, 1988) and stereoselective (Margolin *et al.*, 1991) acylations in organic media. The enzyme also catalyzes peptide synthesis, either by direct reversal of the hydrolytic process or by aminolysis of *N*-protected amino acid or peptide esters (Wong and Wang, 1991). Obtaining high enzyme activity is important for synthetic applications such as these. The *specific* activity of subtilisin E was greatly enhanced by the directed evolution experiment described above. However, the *total* activity of the *Bacillus subtilis* culture supernatant was reduced to <10% that of the starting wild type culture supernatant in the absence of DMF (Table I). This drop in total activity reflects a large decrease in the expression of the 'evolved' 10M enzyme in the *B.subtilis* host. Such changes in expression levels are not unexpected, since DNA base and amino acid substitutions may affect the transcription, translation or enzyme export and

maturation process. This result can be directly attributed to the use of a screening method that focused only on improving specific activity in aqueous DMF.

While mutations which slightly improved specific subtilisin activity in DMF could be found by screening only a few hundred colonies on average, fine-tuning a combination of expression levels and specific activity could be expected to require a search of larger variant libraries. Small libraries of subtilisin variants in *B.subtilis* can be obtained by direct transformation of the ligation products of the PCR random mutagenesis (Chen and Arnold, 1993). The low efficiency of *B.subtilis* transformation with ligated recombinant plasmids, however, limits the size of the mutant library that can be established in this host. Here we report further directed evolution of subtilisin, using a *B.subtilis–Escherichia coli* shuttle vector to facilitate establishment of the mutant library in *B.subtilis*. By slightly altering the screening method to reflect the total subtilisin activity as well as activity in aqueous DMF, significant improvements in both measures were obtained.

## Materials and methods

Restriction enzymes were purchased from Boehringer Mannheim (Indianapolis, IN). Reagents for oligonucleotide synthesis were purchased from Pharmacia (Piscataway, NJ). Succinyl-Ala-Ala-Pro-Phe-*p*-nitroanilide (s-AAPF-pNa) was obtained from Sigma (St Louis, MO). Nitrocellulose membranes were purchased from Schleicher & Schuell (Keene, NH). Biodyne membranes were obtained from Pall (Glen Cove, NY).

*Bacterial strains and plasmids*

*Bacillus subtilis* strain DB428 and *Bacillus* cloning vector pKWZ containing the subtilisin E gene (Park *et al.*, 1989) were kindly provided by R.Doi, University of California, Davis. Plasmid pUC19 and *E.coli* strain ER 1648 were from New England Biolabs (Beverly, MA). Plasmid pUB110 was from Bacillus Genetics Stock Center, Department of Biochemistry, The Ohio State University. Phagemids SKII+ and SKII– were purchased from Stratagene (La Jolla, CA). Competent *E.coli* ER1648 was prepared according to the standard $Ca^{2+}$ protocol (Sambrook *et al.*, 1989). *Bacillus subtilis* DB428 was made competent for transformation as described previously (Chen and Arnold, 1993). *Escherichia coli* ER1648 cells containing plasmid pBE2 were grown on LB agar plates or LB broth supplemented with 20 μg/ml kanamycin. *Bacillus subtilis* DB428 cells harboring plasmids pKWZ or pBE2 were grown on LB agar plates or LB broth supplemented with 50 μg/ml kanamycin.

Table I. Total activities of supernatants of *B.subtilis* expressing wild type and variant subtilisins E

| Subtilisin E | 0% DMF (units/ml) | 20% DMF (units/ml) |
|---|---|---|
| WT | 4.5 | 0.2 |
| 10M | 0.27 | 0.21 |
| 12M | 1.4 | 1.55 |
| 13M | 2.7 | 3.4 |

Assay is for hydrolysis of s-AAPF-pNA in 0.1 mM Tris–HCl, 10 mM $CaCl_2$, pH 8.0 with and without 20% DMF (v/v).

*Construction of B.subtilis–E.coli shuttle expression vector pBE1*

Shuttle vector pBE1 was constructed by ligation of three DNA fragments. The first was the *Sna*BI–*Bam*HI fragment (nucleotides 2164–793) of plasmid pUB110 containing the *Bacillus* replication origin region and the kanamycin resistance gene (McKenzie *et al.*, 1986, 1987). The second and third fragments were generated from pUB110 and pUC19 by the PCR (Saiki *et al.*, 1988) using synthetic oligonucleotide primers. To obtain the second fragment, oligonucleotides MO5B (5′ CTCTTCCTTC GGATCCTATG TAAATCGCTC CTT 3′) and MO3P (5′ CCTCCTTTCC CAGCTGGAAC GAGACTTTGC AGT 3′) were used as the 5′ and 3′ primers, respectively, to amplify the region from nucleotides 1075 to 1651 of pUB110 containing the minus-strand replication origin. The 3′ primer was designed to eliminate the unique *Nde*I site of pUB110 to facilitate the subsequent cloning of the subtilisin E gene for random mutagenesis. The first 10 nucleotides at the 5′ ends of both primers were random sequences to facilitate restriction enzyme digestion. The underlined sequences were the restriction sites for *Bam*HI and *Pvu*II respectively. The third fragment containing the replication origin of *E.coli* plasmid pUC19 (nucleotides 630–1559) was generated using the 5′ and 3′ primers COE5H (5′ TCTTCAGCCTGTTGACCTGC ATTAA-TGAATCGG 3′) and COE3H (5′ TTCTCTTCCTGTTGACT-**AAAAAGAAGCAGGTTTTCTTATACCTGCTTCTTTT-TACAGCTG**ATCTAGGTGAAGATCC 3′), respectively. The first 10 nucleotides of COE5H and COE3H are random sequences. The underlined sequences are restriction sites for *Hinc*II and *Pvu*II respectively. Primer COE3H contains the transcriptional terminator sequence of the subtilisin E gene (shown in bold face), which was designed to be positioned immediately downstream from the kanamycin gene in pBE1. The *Pvu*II site in primer COE3H was used to determine the orientation of the third fragment in pBE1 and to facilitate future plasmid manipulation. The subtilisin E gene was removed from plasmid pKWZ by cutting with *Eco*RI and *Bam*HI and subcloned into *Eco*RI–*Bam*HI-digested pBE1 to form pBE2 (Figure 1).

*PCR-based random mutagenesis*

Random mutagenesis was carried out using error-prone PCR (Leung *et al.*, 1989). Two synthetic oligonucleotides SUB5N (5′ GATCCGAGCGTTGCATATGTGGAAGAAGATCAT 3′) and SUB3B (5′ GGTTCTTGGATCCGATT-CAA CATGCGG-AG 3′) permitted amplification of the full-length subtilisin E sequence. The underlined sequences are *Nde*I and *Bam*HI restriction sites for SUB5N and SUB3B respectively, which allowed the PCR products to be ligated with vector pBE2 digested with the same enzymes. The ligation mixture was then transformed into *E.coli* ER1648 competent cells to generate a library of subtilisin E mutant genes. The mutant library was isolated from *E.coli* ER1648 host cells and transferred into *B.subtilis* DB428 competent cells for expression and screening.

*Screening for enhanced subtilisin activity in aqueous DMF*

Modified Schaefer's agar plates (Leighton and Doi, 1971) containing 1% casein were used for *B.subtilis* growth and subtilisin expression. Agar plates were first covered with two membranes, one organic solvent-resistant Biodyne nylon membrane and one nitrocellulose membrane, onto which *B.subtilis* transformants were plated. *Bacillus subtilis* cells containing an unmutated subtilisin E gene transformed with

pBE2 were used as controls. After overnight incubation at 37°C, observable halos indicated that colonies were secreting active subtilisin. The top membrane carrying the *B.subtilis* colonies was removed and kept at 4°C. The bottom membrane (which contains enzyme at each spot corresponding to individual halos) was then transferred onto a new set of agar plates containing 1% casein and 35–45% DMF. These DMF plates were incubated at 37°C overnight. The halo sizes on both the aqueous and DMF plates were then evaluated and compared with those of unmutated control colonies. Colonies which outperformed the controls were selected and grown in 2 ml modified Schaeffer's medium at 37°C for 48 h in a rotary shaker. The supernatants of the liquid cultures were used for assays of total hydrolytic activity towards the peptide substrate s-AAPF-pNA in 0.1 mM Tris–HCl, pH 8.0 and 10 mM $CaCl_2$. DMF was added as indicated (volume %).

*Sequencing of subtilisin E variant genes*

Selected plasmids were digested with *Eco*RI and *Bam*HI and the subtilisin E gene subcloned into Bluescript SKII+ and SKII– phagemids with *Eco*RI and *Bam*HI. Single-stranded phagemids containing both orientations of a variant subtilisin E gene were prepared according to the supplier's instructions. Dideoxynucleotide DNA sequencing was carried out using the Sequenase II kit (USB).

*Enzyme purification and kinetics*

Subtilisin E purification and kinetic studies were carried out as described previously (Chen and Arnold, 1993). s-AAPF-pNa (0.24 mM) was used to assay hydrolytic activity in 0.1 M Tris–HCl, pH 8.0, 10 mM $CaCl_2$, 37°C, with DMF added as indicated. With the exception of the total activity assays, which were carried out using culture supernatant, all activity assays were carried out with purified enzymes. One unit of subtilisin E is the amount of enzyme required to catalyze the hydrolysis of 1 μmol substrate to products in 1 min.

*Molecular modeling*

Modeling of subtilisin E and its variants was carried out using the InsightII computer program based on the X-ray crystal structure of the subtilisin from *Bacillus mesentericus* (Dauter *et al.*, 1991) (Protein Data Bank reference: 1MEE). Subtilisin E from *B.subtilis* differs from subtilisin from *B.mesentericus* only at three out of 275 amino acid positions. The energy minimized s-AAPF-pNA–subtilisin complex is in good agreement with a similar model prepared for subtilisin BPN' and s-AAPF-pNa (Wells and Estell, 1988).

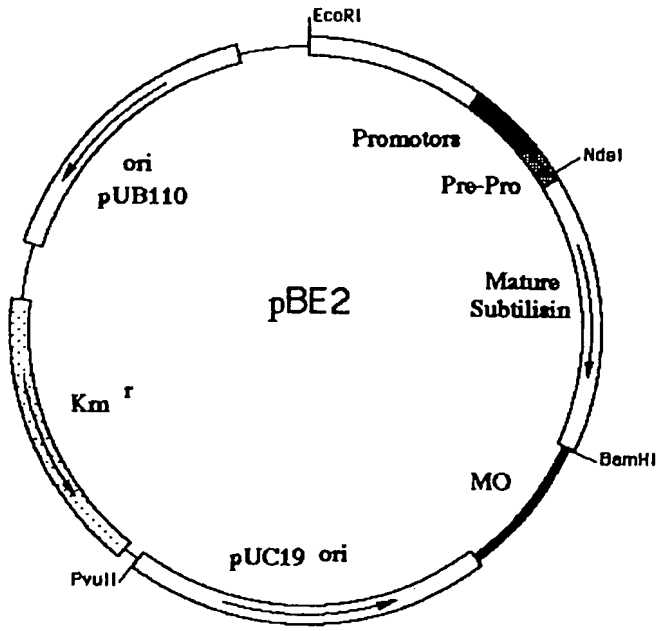



Fig. 1. The *B.subtilis–E.coli* shuttle vector pBE2 carrying a subtilisin E gene, constructed from pUB110 and pUC19 (see Materials and methods). The *Nde*I site of pUB110 downstream from the minus origin (MO) region was eliminated so that the *Nde*I site in the subtilisin E gene can be used for generating the PCR-based mutant library.

## Results and discussion

*The B.subtilis–E.coli shuttle expression vector pBE2 for directed evolution*

Successful application of directed enzyme evolution requires screening large numbers of variants for the desired feature(s). Provided the screen is sufficiently sensitive, mutations conferring small enhancements can be identified and subsequently combined to achieve the desired result. The size of the mutant library that can be established in *E.coli* is generally not a limiting factor for screening, since even the most rapid screening methods can only handle ~$10^6$ or fewer clones. The direct cloning of mutant subtilisin E genes into *B.subtilis* competent cells, however, occurs with low efficiency (a few hundred transformants per microgram of DNA) relative to *E.coli* ($10^6$–$10^7$ transformants/μg DNA). Screening in *B.subtilis* is therefore quickly limited by the size of the mutant library. This is at least partially due to the fact that only multimeric linear or circular forms of plasmids can efficiently transform *B.subtilis* (Canosi *et al.*, 1978; Mottes *et al.*, 1979; de Vos *et al.*, 1981). The monomeric ligation products from the linearized plasmid and target gene cannot transform *B.subtilis* competent cells (de Vos *et al.*, 1981).

The *B.subtilis–E.coli* shuttle vector pBE1 was designed in order to facilitate the establishment of larger mutant libraries in *B.subtilis*. The shuttle vector was constructed by ligating

Table II. Kinetic parameters $k_{cat}$, $K_m$ and $k_{cat}/K_m$ for hydrolysis of S-AAPF-pNn by subtilisin E variants in 0.1 M Tris–HCl, 10 mM $CaCl_2$, pH 8.0, containing 0, 20 and 60% (v/v) DMF at 37°C

| Variant | 0% DMF | | | 20% DMF | | | 60% DMF | |
|---|---|---|---|---|---|---|---|---|
| | $K_m$ (mM) | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_m$ ($M^{-1}s^{-1}\times10^{-3}$) | $K_m$ (mM) | $k_{cat}$ ($s^{-1}$) | $k_{cat}/K_m$ ($M^{-1}s^{-1}\times10^{-3}$) | $k_{cat}/K_m$ ($M^{-1}s^{-1}\times10^{-3}$) | |
| WT | 0.56 | 21 | 38 | 17 | 12 | 1.4 | 0.014 | (1)[a] |
| 10M | 0.1 | 27 | 270 | 0.7 | 73 | 99 | 2.2 | (157) |
| 13M | 0.067 | 39 | 582 | 0.4 | 98 | 245 | 6.6 | (471) |

[a]Catalytic efficiency relative to wild type subtilisin E.

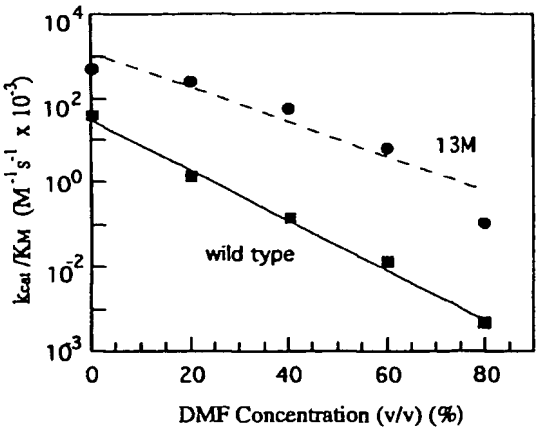



**Fig. 2.** Comparison of the catalytic efficiencies of wild type and 13M subtilisin E towards hydrolysis of s-AAPF-pNa in 0.1 M Tris-HCl, pH 8.0, 10 mM $CaCl_2$ with DMF (v/v), at 37°C.

three DNA fragments from plasmids pUB110 and pUC19 containing the replication origin region and the kanamycin resistance gene of pUB110, the portion of pUB110 covering the minus-strand replication origin and the replication origin region of pUC19 (see Materials and methods). The subtilisin E gene was subcloned from plasmid pKWZ to form pBE2 (Figure 1). The stabilities of plasmids pBE1 and pBE2 were found to be similar to that of pUB110 in *B.subtilis* (data not shown).

With the aid of shuttle vector pBE2, mutant libraries established first in *E.coli* could be easily amplified, isolated by a routine miniprep method and transferred into frozen *B.subtilis* competent cells. A significant fraction of plasmids isolated from *E.coli* ER 1648 is in the form of a supercoiled dimer of multimers (data not shown), which transform *B.subtilis* competent cells with high efficiency (de Vos *et al.*, 1981).

*Screening for increased subtilisin activity in aqueous DMF*

Sequential rounds of random mutagenesis and screening to identify and accumulate positive mutations resulted in a subtilisin E variant (10M) with significantly enhanced specific activity towards peptide hydrolysis in aqueous DMF (Chen and Arnold, 1993). The $k_{cat}/K_m$ of the 10M variant in 60% DMF, for example, is enhanced 157-fold relative to the wild type enzyme for hydrolysis of s-AAPF-pNa (Table II). However, as shown in Table I, the total hydrolytic activity of the *B.subtilis* culture supernatant was reduced to less than 10% that of the starting wildtype culture, reflecting a very significant drop in the overall expression level. This is consistent with our observation that *B.subtilis* colonies expressing the 10M mutant generate much smaller halos on 1% casein plates containing no DMF than colonies expressing the wild type enzyme.

Starting with the gene for the 10M subtilisin E, two additional generations of PCR-based random mutagenesis and screening were conducted, using slightly modified screening criteria. After the first round of mutagenesis, one colony with a larger halo than the control colony (cells containing the parent 10M subtilisin E gene) on both casein plates containing DMF and no DMF was selected from several thousand colonies screened. That gene was isolated and subjected to a second round of random mutagenesis. A single colony producing larger halos on both aqueous and DMF plates was similarly identified from that *B.subtilis* library. The enzymes isolated from these first- and second-generation colonies are termed 12M and 13M, respectively.

**Table III.** Residual activities (reported as % of starting activity) of wild type, 10M and 13M subtilisins E after incubation in 70% DMF at 25°C

| Subtilisin E | Incubation time | | |
|---|---|---|---|
| | 60 h | 130 h | 460 h |
| Wild type | 91 | 88 | 48 |
| 10M | 88 | 85 | 42 |
| 13M | 92 | 89 | 50 |

**Table IV.** DNA and amino acid substitutions in 13M subtilisin E, obtained by site-directed mutagenesis and random mutagenesis and screening for enhanced specific activity in aqueous DMF

| Base | Base substitution | Position in codon | Amino acid | Amino acid substitution |
|---|---|---|---|---|
| 595 | A→C | 3 | 47 | Silent |
| 596 | G→A | 1 | | |
| 597 | C→G | 2 | 48 | Ala→Arg |
| 632[a] | G→A | 1 | 60 | Asp→Asn |
| 643 | T→C | 3 | 63 | Silent |
| 664 | T→G | 3 | 70 | Silent |
| 706 | T→C | 3 | 84 | Silent |
| 744[a] | A→G | 2 | 97 | Asp→Gly |
| 762[a] | A→G | 2 | 103 | Gln→Arg |
| 773 | A→G | 1 | 107 | Ile→Val |
| 838 | A→G | 3 | 128 | Silent |
| 846 | G→A | 2 | 131 | Gly→Asp |
| 871 | T→C | 3 | 139 | Silent |
| 921 | A→G | 2 | 156 | Glu→Gly |
| 946 | A→G | 3 | 164 | Silent |
| 996 | A→G | 2 | 181 | Asn→Ser |
| 998 | A→G | 1 | 182 | Ser→Gly |
| 1016 | T→C | 1 | 188 | Ser→Pro |
| 1033 | T→C | 3 | 193 | Silent |
| 1071 | A→T | 2 | 206 | Gln→Leu |
| 1107[a] | A→G | 2 | 218 | Asn→Ser |
| 1126 | T→C | 3 | 224 | Silent |
| 1141 | A→G | 3 | 229 | Silent |
| 1153 | A→G | 3 | 233 | Silent |
| 1207 | A→G | 3 | 251 | Silent |
| 1216 | A→G | 3 | 254 | Silent |
| 1217 | A→G | 1 | 255 | Thr→Ala |

DNA and amino acid substitutions resulting from two generations of random mutagenesis and screening of 10M subtilisin E (Chen and Arnold, 1993) for enhanced total and specific activity in aqueous DMF are underlined.

[a]Substitutions incorporated by site-directed mutagenesis into DNA coding for 4M subtilisin E used as template for sequential random mutagenesis.

*Catalytic activities of the 'evolved' enzymes*

The total activities of supernatants of *B.subtilis* cultures expressing wild type, 10M and the further-evolved 12M and 13M subtilisin E genes towards hydrolysis of s-AAPF-pNa are summarized in Table I. The 12M variant from the first round of random mutagenesis showed significantly greater activity with and without DMF, relative to its parent 10M. The total activity is in fact slightly greater in the presence of 20% DMF than in purely aqueous buffer. The 13M variant from the second generation continues this trend: the total activity in the absence of DMF is 10-fold greater than the starting 10M variant, while activity in 20% DMF is improved by a factor of 16. In contrast to wild type subtilisin E, which

Fig. 3. Model of subtilisin E showing the 10 amino acid mutations in 10M (yellow) and the additional three mutations identified in the 13M variant (blue) evolved for increased total activity in DMF. Bound $Ca^{2+}$ and peptide substrate s-AAPF-pNA are shown in gray.

retains only 4.4% of its activity in the presence of 20% DMF, 13M subtilisin E is actually 25% more active in 20% DMF than with no DMF at all.

The subtilisin variants were purified and further characterized. The kinetic parameters for the hydrolysis of s-AAPF-pNa by wild type subtilisin E and the selected variants are summarized in Table II. The specific catalytic efficiency $k_{cat}/K_m$ of 13M subtilisin E is approximately 2.4-fold higher than that of 10M in 20% DMF and 175 times greater than that of the wild type enzyme. The increase in catalytic efficiency reflects an increase in $k_{cat}$ as well as a decrease in the $K_m$ (Table II).

As observed previously for the 10M variant, the enhancement in the catalytic efficiency of 13M subtilisin E extends to very high concentrations of DMF (Figure 2). The catalytic efficiency of 13M subtilisin is less sensitive to DMF than the wild type enzyme over the entire range of DMF concentration. In 60% DMF, the $k_{cat}/K_m$ for 13M is 3-fold higher than 10M and 471-fold higher than the wild type subtilisin E.

The loss of activity that the wild type enzyme experiences in DMF has been substantially recovered by sequential generations of random mutagenesis and screening, to the point that the evolved enzyme is actually better at this particular catalytic task in low DMF concentrations (<40%) than the wild type enzyme in purely aqueous media. In 20% DMF, 13M subtilisin E is ~6.5 times more efficient than the wild type enzyme in aqueous solution (Figure 2). In 40% DMF, it is still 1.5 times more efficient than is wild type subtilisin E in the absence of DMF. This result is not as surprising as it might seem at first glance. It should be noted that in aqueous media the evolved enzymes are all 5- to 10-fold more efficient than the wild type enzyme. This improvement in the $k_{cat}/K_m$ is largely due to a decrease in the $K_m$ for the specific substrate s-AAPF-pNa used for screening (Chen and Arnold, 1993). Because the enzyme's evolution has been directed towards improving activity towards this particular substrate as well as catalytic activity in DMF, it is not unreasonable that the evolved enzyme is even more efficient than wild type.

*Enzyme stability*

The time courses of deactivation of the wild type, 10M and 13M subtilisins were compared at 25°C in 70% DMF (Table III). Under these conditions the evolved subtilisin will catalyze the polymerization of amino acid esters (Chen and Arnold, 1993). After 460 h incubation, mutant 13M retained ~50% activity, essentially the same as the wild type enzyme. As observed previously for the directed evolution of the 10M parent enzyme (Chen and Arnold, 1993), the mutations that enhance catalytic efficiency and total expressed activity have no significant net effect on the enzyme's stability, as measured

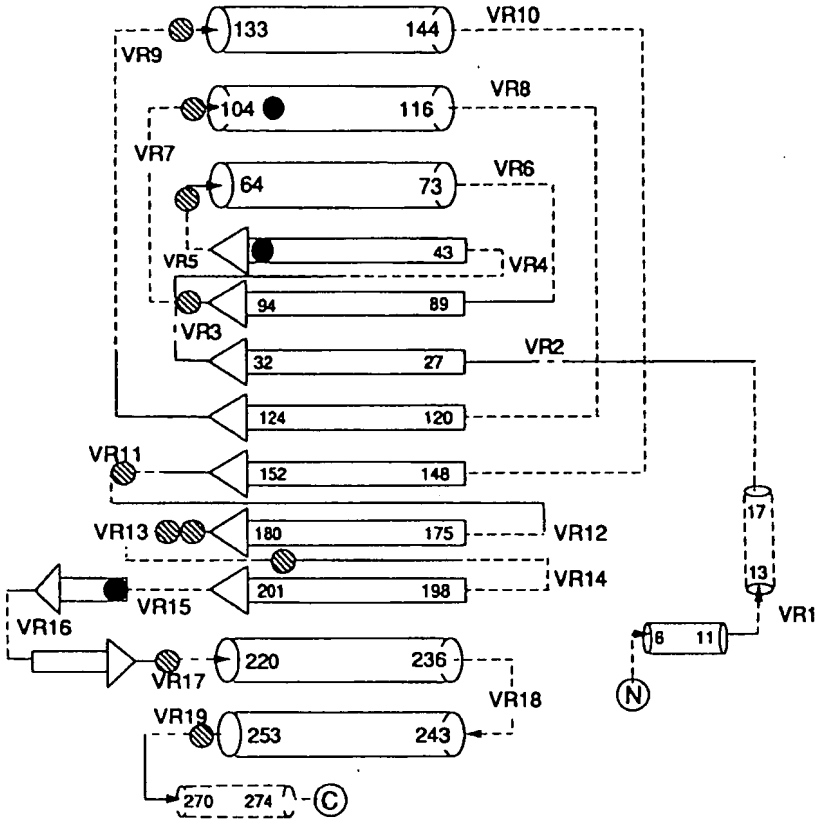



**Fig. 4.** Schematic representation of the secondary structure topology of subtilisin. α-Helices are represented as cylinders and β-sheets as arrows. Solid lines indicate conserved regions in subtilisin sequences. Dashed lines indicate variable regions (VRs) [redrawn from Siezen *et al.* (1991)]. Locations of amino acid substitutions in 10M subtilisin E are indicated by O. The three additional substitutions in 13M are indicated by ●.

by residual activity after incubation in aqueous DMF. Since the enzyme is sufficiently stable to maintain its activity during the screening process, in the presence of DMF, obviously no benefit is derived from mutations that only enhance stability. The fact that mutations enhancing catalytic efficiency occur without corresponding changes in stability indicate that activity and stability in DMF are not coupled traits.

*DNA and amino acid sequence changes*

Site-directed mutagenesis followed by five generations of sequential random mutagenesis and screening resulted in the accumulation of a total of 27 DNA base substitutions coding for 13 amino acid substitutions in 13M subtilisin E (Table IV). Two generations of random mutagenesis and screening resulted in seven new base changes (underlined in Table IV) distributed across the 800 bp portion of the 10M subtilisin E gene sequence subjected to random mutagenesis. The PCR mutagenesis method used previously to obtain 10M subtilisin E involved introduction of dGTP and dATP in a 10:1 ratio and resulted in a preponderance of transitions (A→G or T→C). A more balanced distribution of transversions (three) and transitions (four) was obtained using the method of Leung *et al.* (1989). These seven base changes give rise to three amino acid substitutions, two of which (A48R and I107V) appeared during the first cycle of mutagenesis and screening (12M subtilisin E) and a third (Q206L) in the second cycle (13M). Substitutions I107V and Q206L are the results of single base substitutions in their corresponding codons. Substitution of Ala at position 48 by Arg, on the other hand, is the result of a double base change in the codon for amino acid 48.

As shown in Figure 3, the 10 amino acid substitutions in the parent 10M enzyme (shown in yellow) are clustered on the face of enzyme that harbors the active site and substrate binding pocket. The three new amino acid substitutions found in 13M subtilisin E (shown in blue) also appear more or less on the same face. All three new substitutions reside on the enzyme surface, as do all the previous 10 amino acid substitutions. Figure 3 shows that the 10 amino acid substitutions in 10M subtilisin E are located in loop structures connecting elements of secondary structure. In contrast, the three new amino acid substitutions in 13M are located in α-helix (I107V) or β-sheet (A48R and Q206L) secondary structures. I107V lies in the substrate binding pocket, where it interacts directly with the peptide substrate (shown in gray), while the remaining two mutations are located well outside this area. The substitution of Ile107 with Val was found in an earlier study to increase the stability slightly of subtilisin from *Bacillus amyloliquefaciens* towards autolysis under alkaline conditions (Cunningham and Wells, 1987).

Comparison of the amino acid sequences of more than 40 subtilisins from different sources shows a conserved core with insertions and deletions that are preferentially confined to surface loops (Siezen *et al.*, 1991). As indicated in Figure 4, the core α-helix and β-sheet secondary structures are relatively conserved, while the sequences of the peptide loops that connect these secondary structure elements are variable among

the different subtilisins (VRs). Figure 4 clearly shows the positions of the 10 amino acid substitutions in 10M subtilisin E in these variable loops on the surface of the enzyme surrounding the active site. In contrast, the three new amino acid substitutions in 13M lie within secondary structures that are more conserved throughout the different subtilisins.

Although these secondary structures are more conserved in their sequences, variations do occur among subtilisins from different sources. Using the subtilisin sequences compared by Siezen *et al.* (1991), it can be seen that all three of the new amino acids in 13M are found in subtilisins from cyanobacteria (Arg48) and Gram-positive (Val107 and Leu206) or Gram-negative bacteria (Leu206). One subtilisin (from *Staphylococcus epidermis*) contains two of the three new amino acids in 13M (Val107 and Leu206); none of the sequences compared, however, have all three. Thus the individual 'solutions' found by this directed evolution approach have indeed been explored by nature, albeit for other purposes. As noted previously (Chen and Arnold, 1993), at least seven of the 10 amino acid substitutions in 10M subtilisin E are found in subtilisins from other sources. The new 13M subtilisin variant can therefore be thought of as largely a novel 'mix and match' of other, naturally occurring enzymes.

It is unknown how the mutations in the subtilisin gene reduce the expression of active enzyme in the 10M variant or why expression has been recovered in the 13M variant. Mutations at the DNA level can affect the transcription and translation processes, while mutations at the protein level can affect the folding and processing of the enzyme. For example, the subtilisin maturation process requires a pro-sequence which acts as an intramolecular chaperon for the correct folding of the enzyme (Ikemura *et al.*, 1987). Changes in local secondary or tertiary structures resulting from amino acid substitutions could interfere with the maturation process and reduce enzyme yield. The enzyme yield was (purposefully) not maintained during the earlier directed evolution for enhanced specific activity in DMF. The amino acid 'solutions' to this problem found in 10M subtilisin E all share the following features: they are located on the surface of the enzyme around the active site (although not necessarily adjacent) and they are all located in sequence-variable loops. When the screening criteria were modified in the current experiment to include an enhanced enzyme yield, the nature of the 'solutions' changed. That the three new amino acid substitutions in conserved secondary structures in 13M contrast with the positions of the 10 previous mutations in variable loops (Figure 4) hints at a mechanism for recovering total activity that occurs at the protein level.

A basic rule of directed evolution (and any random mutagenesis experiment) is, 'you get what you screen for'. The screen (or selection) should therefore reflect the desired result as closely as possible. In particular, the standards for selecting a positive should include all features of interest, in order to minimize unwanted side-effects such as a lower expression level or changes in other features markedly affected by mutations at the DNA or amino acid levels. These realities often work against the use of a selection scheme that offers the ability to search larger fractions of sequence space for something no-one wants.

Although screening limits the search for beneficial mutations to a very small fraction of sequence space, a workable strategy for directed evolution is to accumulate mutations conferring small enhancements in order to achieve the desired property (Arnold, in press). Sequential generations of error-prone PCR mutagenesis and screening is one effective method by which this can be accomplished. The accumulation of amino acid substitutions in an enzyme can result in significant improvements in performance, particularly for features not optimized under selective pressure.

## Acknowledgements

The authors are grateful to Jeffrey Moore and Huimin Zhao for their assistance with the modeling and Figure 3. This work was supported by the US Department of Energy's program in Biological and Chemical Technologies Research within the Office of Industrial Technologies, Energy Efficiency and Renewables.

## References

Arnold,F.H. (1993) *FASEB J.*, **7**, 744–749.
Arnold,F.H. (1996) *Chem. Eng. Sci.*, in press.
Bryan,P.N., Rollence,M.L., Pantoliano,M.W., Wood,J., Finzel,B.C., Gilliland,G.L., Howard,A.J. and Poulos,T.L. (1986) *Proteins: Struct. Funct. Genet.*, **1**, 326–334.
Canosi,U., Morelli,G. and Trautner,T.A. (1978) *Mol. Gen. Genet.*, **166**, 259–267.
Chen,K.Q. and Arnold,F.H. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 5618–5622.
Cunningham,B.C. and Wells,J.A. (1987) *Protein Engng*, **1**, 319–325.
Dauter,Z., Betzel,C., Genov,N., Pipon,N. and Wilson,K.S. (1991) *Acta Crystallogr.*, **B47**, 707–730.
de Vos,W.M., Venema,G., Canosi,U. and Trautner,T.A. (1981) *Mol. Gen. Genet.*, **181**, 424–433.
Graham,L.D., Haggett,K.D., Jennings,P.A., Le Brocque,D.S. and Whittaker,R.G. (1993) *Biochemistry*, **32**, 6250–6258.
Ikemura,H., Takagi,H. and Inouye,M. (1987) *J. Biol. Chem.*, **262**, 7859–7864.
Joyet,P., Declerck,N. and Gailardin,C. (1992) *Bio/Technology*, **10**, 1579–1583.
Leighton,T.J. and Doi,R.H. (1971) *J. Biol. Chem.*, **246**, 3189–3195.
Leung,D.W., Chen,E. and Goeddel,D.V. (1989) *Technique*, **1**, 11–15.
Liao,H., McKenzie,T. and Hageman,R. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 576–580.
McKenzie,T., Hoshiro,T., Tanaka,T. and Sueka,N. (1986) *Plasmid*, **15**, 93–103.
McKenzie,T., Hoshiro,T., Tanaka,T. and Sueoka,N. (1987) *Plasmid*, **17**, 83–85.
Margolin,A.L., Fitzpatrick,P.A., Dubin,P.L. and Klibanov,A.M. (1991) *J. Am. Chem. Soc.*, **113**, 4693–4694.
Mottes,M., Grandi,G., Sgaramella,V., Canosi,U., Morelli,G. and Trautner,T.A. (1979) *Mol. Gen. Genet.*, **174**, 281–286.
Oliphant,A.R. and Struhl,K. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9094–9098.
Park,S.-S., Wong,S.-L., Wang,L.-F. and Doi,R.H. (1989) *J. Bacteriol.*, **6**, 521–527.
Riva,S. and Klibanov,A.M. (1988) *J. Am. Chem. Soc.*, **10**, 3291–3295.
Riva,S., Chopineau,J., Kieboom,A.P.G. and Klibanov,A.M. (1988) *J. Am. Chem. Soc..*, **110**, 584–589.
Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B. and Erlich,H.A. (1988) *Science*, **239**, 487–491.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Siezen,R.J., De Vos,W.M., Leunissen,J.A.M. and Dijkstra,B.W. (1991) *Protein Engng.*, **4**, 719–737.
Stemmer,W.P.C. (1994) *Nature*, **370**, 389–391.
Strausberg,S.L., Alexander,P.A., Gallagher,D.T., Gilliland,G.L., Barnett,B.L. and Bryan,P.N. (1995) *Bio/Technology*, **13**, 669–673.
Wells,J.A. and Estell,D.A. (1988) *Trends Biochem. Sci.*, **13**, 291–297.
Wong,C.-H. and Wang,K.-T. (1991) *Experientia*, **47**, 1123–1129.

*Received September 5, 1995; accepted on October 14, 1995*

Šič, D. 1990. Regulation o
d. – Acta Hortic. 280: 169

Stimulatory effects of
and protoplast culture

. 1993. Jasmonic acid stim
of garlic in vitro. – J. Plant

The influence of jasmon
er ion uptake of the pota
). – Acta Pharm. 45: 221–

einbothe, C. 1994. JIPs and
ıe expression by jasmonates
ıes and pathogens. – Plant

The biochemistry and
ons of jasmonates. – Annu
ol. 44: 569–589.
. & Koda, Y. 1994. Expan
) jasmonic acid. – Plant Sci

ıorescein diacetate and pro
ility of cultured plant cell

PHYSIOLOGIA PLANTARUM 97: 665–673. 1996
Printed in Denmark – all rights reserved


ISSN 0031-9317

# BT1, a protein critical for in vivo starch accumulation in maize endosperm, is not detected in maize endosperm suspension cultures

Heping Cao and Jack C. Shannon

Cao, H. and Shannon, J. C. 1996. BT1, a protein critical for in vivo starch accumulation in maize endosperm, is not detected in maize endosperm suspension cultures. – Physiol. Plant. 97: 665–673.

Maize *Bt1* is a structural gene for the major 39–44-kDa amyloplast membrane polypeptides. Loss of BT1 in the *brittle1* (*bt1*) mutant results in an 80% reduction in kernel starch. As an initial step toward the understanding of BT1 function, we investigated the relation of BT1 to starch accumulation in the two well-characterized maize endosperm suspension-cultured cell lines which were derived from 10 days post pollination (DPP) kernels of inbred A636 and 12-DPP kernels of the *waxy* mutant in the A636 background. Starch in A636 endosperm cultures accounted for about 1.5% of the fresh weight of cells but BT1 was not detectable in amyloplast membranes or in microsomal membranes isolated from the cultured cells. *Bt1* transcripts were detected in 10- and 20-DPP A636 kernels, but only a trace of *Bt1* transcripts was detected in the suspension-cultured cells. Southern blotting indicated that the *Bt1* gene was present in the genome of the cultured cells. The *Bt1* gene products in A636 endosperm cultures were not increased by treatments which enhance starch accumulation such as sucrose supplementation, chlorocholine chloride (CCC) addition to the medium, or both. A similar pattern of differential *Bt1* gene expression was found in the *waxy* endosperm cultures. These results indicate that starch accumulation in the cultured cells is not correlated with *Bt1* gene expression. The significance of these findings is discussed.

*Key words* – Amyloplast membrane, *brittle1* gene, differential gene expression, endosperm suspension culture, maize, starch synthesis, *Zea mays*.

*H. Cao (present address: Dept of Chemistry and Biochemistry, Texas Tech Univ., Lubbock, TX 79409, USA) and J. C. Shannon (corresponding author, e-mail jack_shannon@agcs.psu.edu), Plant Physiology Program/Dept of Horticulture, Pennsylvania State Univ., University Park, PA 16802, USA.*

## Introduction

Plant plastids are a group of subcellular organelles including chloroplasts, chromoplasts, amyloplasts, etioplasts, leucoplasts and proplastids. Plastids contain a double-membrane envelope, stroma, and often internal membranes as well as one or more starch granules (Boyer et al. 1989). The envelope membrane proteins are essential in the sorting and translocation of nuclear-encoded proteins directed to plastids (Keegstra et al. 1989, Schnell et al. 1994) and in the transport of various metabolites (Heldt et al. 1991). At present, the transport processes and starch biosynthetic pathway(s) in nonphotosynthetic tissues are controversial (Boyer et al. 1989, Okita 1992, Villand and Kleczkowski 1994). The use of maize mutants in the study of starch biosynthesis has contributed much to an improved understanding of the in vivo mechanism of starch synthesis (Nelson and Pan 1995). The study of mutants including *shrunken2* (*sh2*), *brittle2* (*bt2*), *amylose extender* (*ae*) and *waxy* (*wx*) has helped to establish some of the critical steps in starch synthesis (Hannah et al. 1993, Nelson and Pan 1995, Preiss 1991).

Received 27 November, 1995; revised 1 March, 1996

Kernels of the maize starch-deficient mutant *brittle1* (*bt1*), first reported in 1926 (Mangelsdorf 1926, Wentz 1926), accumulate about 20% as much starch as kernels of the normal genotype and sucrose content is increased by more than 500% (Tobias et al. 1992). Several lines of evidence support the conclusion that Brittle1 protein (BT1) is an amyloplast membrane polypeptide and possibly an adenylate translocator. First, the protein sequence deduced from a *Bt1* cDNA has a transit peptide at its amino terminus and a sequence similarity to several mitochondrial inner-envelope translocator proteins (Sullivan et al. 1991). Second, Li et al. (1992) showed that in vitro synthesized BT1 is imported into chloroplast, processed into mature protein, and localized to the inner envelope membranes. Third, BT1 was shown to be localized to maize amyloplast envelope membranes by immunocytolocalization (Sullivan and Kaneko 1995) and by immunoblotting of polypeptides from isolated amyloplast membranes (Cao et al. 1995). BT1 is a cluster of four 39- to 44-kDa amyloplast-specific and integral membrane polypeptides and in normal kernels these polypeptides account for about 40% of the total amyloplast membrane protein (Cao et al. 1995). Fourth, developing *bt1* kernels accumulated over 13-fold higher levels of adenosine 5' diphosphate glucose (ADP-Glc) than normal (Shannon et al. 1996). Finally, amyloplasts isolated from young kernels of the *bt1* mutant were only 25% as active in the uptake of ADP-Glc and conversion to starch as amyloplasts from normal and several other maize endosperm mutants (Liu et al. 1992).

For several decades, in vitro cultures of maize endosperm cells grown in chemically defined media have been suggested as model systems for in vivo studies of starch biosynthesis (Chu and Shannon 1975, Shannon 1982). The endosperm culture system has been characterized for cell growth (Chu and Shannon 1975), cell friability and ultrastructure (Felker 1987, Felker et al. 1989), sugar uptake (Felker and Goodwin 1988), and starch characteristics (Saravitz and Boyer 1987). A transient accumulation and mobilization of starch occur during a normal 7- or 14-day culture cycle. The cultured cells accumulate starch to about 5–10% of their dry weight at mid culture cycle and then it declines to a low level prior to subculture (Chu and Shannon 1975). Chu and Shannon (1975) also reported that starch content was increased up to 12 to 14% of the dry weight by supplementation of the cultures several days after subculture with additional sucrose. Thus, clearly these cultured maize endosperm cells are capable of starch biosynthesis and we were therefore interested in using them to study BT1 function. As an initial step, we investigated the relation of BT1 to starch accumulation in the two well-characterized maize endosperm suspension-cultured cell lines initially derived from 10 days post pollination (DPP) endosperm of the normal A636 inbred, and from the endosperm of 12-DPP *wx* mutant kernels. Both suspension-cultured cell lines, A636 and *wx*, contained the *Bt1* gene (as detected by Southern blotting), but only traces of *Bt1* transcripts were detected by northern analysis. The presence of BT1 polypeptide was below the level of detection by immunoblot assay. Our results indicate that starch accumulation in the cultured cells is not correlated with the presence of BT1 in the amyloplast membrane.

*Abbreviations* – ADP-Glc, ADP-glucose; AGPase, ADP-Glc pyrophosphorylase; *ae*, *amylose extender*; *bt1*, *brittle1*; *Bt1*, *Brittle1* gene; BT1, Brittle1 protein or Brittle1 polypeptide; CCC, chlorocholine chloride; DPP, days post pollination; PMSF, phenylmethylsulfonyl fluoride; *wx*, *waxy*.

## Materials and methods

### Plants, chemicals and manipulation

Maize (*Zea mays* L. inbred A636 and the endosperm mutant, *waxy*, in a near isogenic A636 background) plants were grown in a greenhouse as described previously (Tobias et al. 1992). The plants were self- or sib-pollinated. Ears were quick-frozen in liquid nitrogen and stored at –70°C until used, except for fresh ears that were used to isolate amyloplasts and amyloplast membranes. The common reagents were purchased from either Sigma (St Louis, MO, USA) or Fisher Scientific (Pittsburgh, PA, USA). Special reagents and kits are indicated in the text. General molecular biological techniques followed the standard procedures (Ausubel et al. 1993, Sambrook et al. 1989). Manipulations of membranes and proteins were the same as those previously described (Cao et al. 1995).

### Maize endosperm suspension cultures

Maize endosperm suspension cultures were derived from 10-DPP endosperm of inbred A636 and 12-DPP endosperm of the *wx* mutant in a near isogenic A636 background according to the procedure of Shannon (1982). Prior to this study, the endosperm suspension cultures of A636 and *wx* had been in culture for about 7 and 3 years, respectively. The suspension cultures were routinely grown in 250-ml Erlenmeyer flasks containing 80 ml of medium devoid of added growth regulators according to the protocol of Shannon (1982). To enhance starch accumulation in some experiments, cells were subcultured into the standard medium (80 ml) supplemented with 10 ml of a filter-sterilized chlorocholine chloride (CCC) stock to yield 90 ml of medium containing 100 μ*M* CCC or cells subcultured to the standard medium were supplemented 3 days after subculture with 10 ml of a sterile solution containing 8 g of sucrose (Chu and Shannon 1975). In the combination treatment, cells subcultured into the CCC-containing medium were also supplemented with sucrose as above (final volume = 100 ml). All cultures were harvested 7 days after subculture (4 days after sucrose supplementation) by vacuum filtration and washed twice with distilled water. Fresh cells were used to isolate amyloplast membranes. DNA, RNA, and microsomal membranes were isolated from frozen cells

### traction and determinati
### lysaccharides

lls (2 g of fresh weig
olytron (Brinkmann Ins
imes for 10 s each in 1
o) chloroform:water, 1:
as centrifuged at 800
ashed 3 times by susp
ntrifugation as above.
ended in 15 ml of 10%
and then centrifuged at
as reextracted 3 times l
hanol and centrifuged
ractions were done in th
egarded as total 10%
ides (total polysacchari
ed water by heating for
the phenol-$H_2SO_4$ proce
used to estimate total po
ame starch content, an
olysaccharide fraction
with 5 units $ml^{-1}$ of amy
odium acetate, pH 4.8.
ng water bath for 1 min
min and the quantity of
elson's reducing sug
0.2). Structural polysa
the difference betw
arch.

### Determination of α-amylas

fresh cells were homog
for 10 s each in 3 volum
in 100 m*M* potassium
m*M* ethylenediaminetet
m*M* dithioerythritol. Fo
ayer of Miracloth and c
min the supernatant wa
ctivity by the starch a
1983).

### troduction of protoplasts

Before isolating amylop
dosperm suspension-cu
roduced from the cultu
ninecultured cells deriv
weight) were plasmolyz
digestion medium: 50
ulfonic acid [MES], pH
*M* EDTA, 0.5 *M* sorbi
digested in 2 ml of 1
100 U $ml^{-1}$ cellulase (S
*riniculosum*) and 10 U
from *Aspergillus niger*)
digestion of the cells w

etected by northern
lypeptide was belo
ot assay. Our results
n the cultured cells
f BT1 in the amylo

glucose; AGPase, AD
*extender*; *bt1*, *brittle*
in or Brittle1 polypep
days post pollination;
*waxy*.

ion

A636 and the endosp
genic A636 backgro
nouse as described
e plants were self- o
ozen in liquid nitroge
except for fresh ear
sts and amyloplast
were purchased from
JSA) or Fisher Scien
l reagents and kits a
nolecular biological
procedures (Ausube
. Manipulations of
same as those previ

ltures

cultures were derive
d A636 and 12-DP
near isogenic A636
edure of Shannon
erm suspension cultu
re for about 7 and 3
cultures were rou
flasks containing 80
vth regulators accord
). To enhance starch
ts, cells were subcul
ml) supplemented w
rocholine chloride (C
n containing 100 μ*M*
dard medium were
e with 10 ml of a ster
crose (Chu and Sha
eatment, cells subcul
edium were also
e (final volume = 10
7 days after subcul
ation) by vacuum fil
led water. Fresh cel
mbranes. DNA, RN
isolated from frozen

**Extraction and determination of starch and structural polysaccharides**

cells (2 g of fresh weight) were homogenized with a Polytron (Brinkmann Instruments, Westbury, NY, USA) times for 10 s each in 15 ml of ice-cold MCW (methanol:chloroform:water, 13:4:3, v/v/v). The homogenate st centrifuged at 800 *g* for 10 min. The pellet was ashed 3 times by suspension in 10 ml of MCW and ntrifugation as above. The washed pellet was susended in 15 ml of 10% ethanol, held overnight at 4°C nd then centrifuged at 3 500 *g* for 20 min. The pellet as reextracted 3 times by suspension in 10 ml of 10% thanol and centrifuged at 3 500 *g* for 20 min. All exactions were done in the cold (~4°C). The final pellet, garded as total 10% ethanol-insoluble polysaccharides (total polysaccharides), was suspended in deionized water by heating for 15 min in a boiling water bath. The phenol-$H_2SO_4$ procedure (Dubois et al. 1956) was used to estimate total polysaccharide content. To determine starch content, an aliquot of the suspended total polysaccharide fraction was digested overnight at 37°C with 5 units $ml^{-1}$ of amyloglucosidase (Sigma) in 0.5 *M* odium acetate, pH 4.8. The digest was heated in a boiling water bath for 1 min, centrifuged at 20 000 *g* for 10 min, and the quantity of glucose released determined by Nelson's reducing sugar test (Hodge and Hofreiter 962). Structural polysaccharide content was calculated s the difference between total polysaccharides and tarch.

**Determination of α-amylase activity**

Fresh cells were homogenized with a Polytron 4 times for 10 s each in 3 volumes of extraction buffer containing 100 m*M* potassium phosphate buffer (pH 7.0), 5 m*M* ethylenediaminetetraacetic acid (EDTA), and 1 m*M* dithioerythritol. Following filtration through one layer of Miracloth and centrifugation at 3 500 *g* for 30 min, the supernatant was used to determine α-amylase activity by the starch azure assay (Doehlert and Duke 1988).

**Production of protoplasts**

Before isolating amyloplast membranes from maize endosperm suspension-cultured cells, protoplasts were produced from the cultured cells as follows. The suspension-cultured cells derived from A636 (100 g of fresh weight) were plasmolyzed in 2× digestion medium (1× digestion medium: 50 m*M* 2-[*N*-morpholino]ethanesulfonic acid [MES], pH 5.6, 1 m*M* KCl, 1 m*M* $MgCl_2$, 1 m*M* EDTA, 0.5 *M* sorbitol) at 30°C for 30 min and then digested in 2 ml of 1× digestion medium containing 100 U $ml^{-1}$ cellulase (Sigma, C-0901 from *Penicillium funiculosum*) and 10 U $ml^{-1}$ pectinase (Sigma, P-9932 from *Aspergillus niger*) per gram fresh weight of cells. Digestion of the cells was carried out in a 3.5-l flat-bottomed glass vessel (20 cm in diameter). After 2 h at 30°C with gentle agitation, release of protoplasts was complete and the vessel was inclined and protoplasts and debris allowed to settle at 1 *g* for 30 min. The supernatant was carefully syphoned off and replaced with half original volume of digestion medium less enzymes. The resulting suspension was then filtered through Miracloth and spun at 14 *g* for 15 min. The supernatant was removed and protoplast resuspended in 1 ml of homogenization buffer (50 m*M* *N*-[2-hydroxyethyl]piperazine-*N*'-[2-ethanesulfonic acid] [HEPES], pH 7.5, 10 m*M* KCl, 1 m*M* $MgCl_2$, 1 m*M* EDTA, 1 m*M* dithiothreitol [DTT], 1 m*M* phenylmethylsulfonyl fluoride [PMSF], and 0.5 *M* sorbitol) minus PMSF and DTT per gram of cells and centrifuged as above. The washed and pelleted protoplasts were suspended in 2 volumes of homogenization buffer.

**Isolation of amyloplast membranes and marker enzyme analysis**

Amyloplast membranes from developing kernels were isolated as described previously (Cao et al. 1995). Amyloplast membranes from A636 endosperm cultures were isolated from the purified protoplasts as follows. Protoplasts were ruptured either by homogenization for 1 s in a VirTis 23 homogenizer (The VirTis Company, Gardiner, NY, USA) operated at maximum power as described by Cao et al. (1995), or by two passages through a nylon mesh (10 μm opening) positioned at the lower end of a 50-ml disposable plastic syringe as described by Shannon et al. (1987). The homogenate or lysate was passed through one layer of 10-μm pore-sized nylon. The filtrate was centrifuged at 100 *g* for 15 min to pellet the amyloplasts. The amyloplasts were suspended in homogenization buffer and purified through a Percoll density gradient consisting of 20, 30, and 40% of Percoll in homogenization buffer minus PMSF and DTT. Amyloplasts in the 20/30% and 30/40% interfaces and pellet layer were collected. The purity of amyloplasts was determined by marker enzyme analyses using the commonly used marker enzymes, i.e. alcohol dehydrogenase (cytosol), cytochrome *c* oxidase (mitochondria), catalase (microbodies), cyanide-insensitive NADH-cytochrome *c* reductase (ER) as described previously (Cao et al. 1995). Amyloplast membranes were isolated from the purified amyloplasts using the same procedure as that used for developing endosperm (Cao et al. 1995).

**Isolation of microsomal membranes**

The procedure for isolating microsomal membranes from developing kernels was described previously (Cao et al. 1995). A similar procedure was used to isolate microsomal membranes from suspension-cultured endosperm cells. The microsomal membrane pellet was suspended in TES buffer (10 m*M* Tricine, pH 7.2, 1 m*M* EDTA, 0.2 *M* sucrose).

### Protein determination, SDS-PAGE and western blotting

Protein content was determined as described (Cao et al. 1995) with the Bio-Rad protein assay (Bio-Rad Laboratories, Richmond, CA, USA) and using bovine serum albumine (BSA) as a standard. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) followed that of Sambrook et al. (1989). The western blotting procedure used was detailed previously (Cao et al. 1995). Briefly, after electrophoresis, the polypeptides in the gel were electroblotted onto a piece of nitrocellulose membrane and incubated sequentially with BT1 antibody and with the secondary antibody, goat anti-rabbit IgG-alkaline phosphatase conjugate (Immun-Blot Assay Kit, Bio-Rad Laboratories). Finally, the target polypeptide(s)-primary and secondary antibody conjugates were detected using the color reaction agent, 5-bromo-4-chloro-3-indolyl phosphate and nitro blue tetrazolium chloride (Immun-Blot Assay Kit, Bio-Rad Laboratories). The primary antibodies were raised against a fusion protein formed from 56 amino acids of the carboxyl terminus encoded by the *Bt1* cDNA and glutathione-*S*-transferase (Sullivan and Kaneko 1995). Each western blot was repeated at least 3 times.

### RNA isolation and northern blotting

Total RNA was isolated from the endosperm of developing kernels and suspension-cultured endosperm cells by the hot phenol procedure of de Vries et al. (1988). The pBt1cDNA1.7 plasmid (Sullivan et al. 1991) was transformed into *E. coli* bacteria (XL1) by the rapid freezing method (Takahashi et al. 1992) and produced by the method of Lee and Rasheed (1990). The *Bt1* cDNA probe was prepared from the purified *Bt1* cDNA inserts using the Random Primed DNA Labeling Kit (Boehringer Mannheim Biochemicals, Indianapolis, IN, USA). RNA was separated in 1% agarose gel in 3-(*N*-morpholino)propanesulfonic acid (MOPS) buffer (Fourney et al. 1990). The RNA integrity and equal-loading of the samples were confirmed by examining band sharpness and staining intensity of the 28S and 18S rRNA bands on the ethidium-bromide-stained gel (Ausubel et al. 1993, Cao et al. 1996). The RNA was then transferred to a piece of nylon membrane (Hybond™-N, Amersham International plc, Amersham, UK) by capillary blotting (Sambrook et al. 1989). Following UV crosslinking, the blot was prehybridized for 1 to 2 h at 65°C in a buffer containing 5× SSPE (1× SSPE: 0.18 *M* NaCl, 10 m*M* sodium phosphate, pH 7.7, 1 m*M* EDTA) (Ausubel et al. 1993), 5× Denhardt's solution (Ausubel et al. 1993), and 0.5% (w/v) SDS, and hybridized for 16 to 20 h at 65°C. After hybridization, the blot was washed twice for 10 min each at room temperature in 2× SSPE containing 0.1% SDS, once for 15 min at 65°C in 1× SSPE containing 0.1% SDS solution, and finally for 15 min at 65°C in 0.1× SSPE containing 0.1% SDS before autoradiography. Northern hybridizations were repeated at least 3 times.

### DNA isolation and Southern blotting

DNA was isolated from developing kernels and from suspension-cultured endosperm cells following the protocol of Junghans and Metzlaff (1990). Following *Eco*R I digestion, DNA fragments were separated on 0.7% agarose gel in TAE buffer (40 m*M* Tris-acetate, 1 m*M* EDTA), and transferred to a piece of Nylon-N⁺ membrane with 0.4 *M* NaOH for 18 to 20 h by capillary transfer (Sambrook et al. 1989). The blot was prehybridized at 65°C for 4 h in a buffer containing 0.5 *M* $Na_2HPO_4$, pH 7.2, 7% SDS, 100 μg $ml^{-1}$ salmon sperm DNA (Sigma), and then hybridized at 65°C for 16 to 20 h in the same buffer plus freshly denatured *Bt1* cDNA probe prepared as above for northern blotting. After hybridization, the blot was washed at room temperature for 5 min in 2× SSC (1× SSC: 0.15 *M* NaCl, 15 m*M* sodium citrate, pH 7.0) (Ausubel et al. 1993) plus 0.5% SDS, 15 min in 1× SSC plus 0.1% SDS, 30 min in 0.1× SSC plus 0.5% SDS, and finally at 65°C for 30 min in 0.1× SSC plus 0.5% SDS before autoradiography.

## Results

### Characterization of starch accumulation in A636 endosperm cultures

Before we examined *Bt1* gene expression in maize endosperm suspension-cultured cells, we characterized starch accumulation in the A636 derived endosperm cultures. After a 7-day culture cycle, the fresh weight of the endosperm cultured cells increased from approximately 2.1 g to 11.2 g (80 ml suspension)$^{-1}$ (Tab. 1). The

polysaccharides made
eight, with starch ac
olysaccharides (Tab. 1)
bberellic acid ($GA_3$) b
aused an increase in fre
and structural polysacch
ivity remained essential
sults indicate that A636
ecient for starch biosynth

olation of amyloplast me
ultures

We recently reported th
olypeptides specific t
branes (Cao et al. 1995
membranes were isolat
ured endosperm cells to
membranes of in vitro c
Amyloplasts in these cul
ormally developing ma
myloplasts in the pellet
plasts were also located
and 30/40% Percoll laye
dient purification. Thus
the two interfaces were
Marker enzyme ana
combined "purified amy
cytosol marker, alcohol
ghtly contaminated by
ytochrome *c* oxidase a
NADH-cytochrome *c* re
(4.4% of the total cat
Following removal of
ractions, the membrane
ed amyloplasts were ye
amyloplast envelope
al. 1995), potato tubers (
ycamore suspension-cu
Harinasut et al. 1988). I
membrane fraction" wa
plasts it is concluded t
ghly enriched in amylo

Tab. 2. Purity of amyloplasts
cribed in Materials and met
hate), in the first 100 *g* cen
All samples were suspended
ivities were determined. D
eight of endosperm cells as

| Preparation | P |
|---|---|
| Homogenate | |
| Crude amyloplast | 18. |
| Purified amyloplast | 1. |
| Compartment | |

Tab. 1. Starch accumulation in cells from A636 endosperm suspension cultures. A636 endosperm cell cultures were subcultured into standard medium with or without CCC as described in Materials and methods. Cultures inoculated at a start density of approximately 2 flask$^{-1}$ were grown for 7 days. Data are the average of two flasks plus or minus the SE.

| Measurement | Control | + CCC | % of Control |
|---|---|---|---|
| Fresh weight (g flask$^{-1}$) | 11.2 ± 0 | 13.0 ± 0.2 | 116.1 |
| Total polysaccharides (mg flask$^{-1}$) | 274.1 ± 18.2 | 376.1 ± 28.8 | 137.2 |
| Starch (mg flask$^{-1}$) | 169.8 ± 11.9 | 225.1 ± 14.8 | 144.8 |
| Structural polysaccharides (mg flask$^{-1}$) | 104.3 ± 6.3 | 151.0 ± 14.0 | 132.6 |
| α-Amylase (μmol min$^{-1}$ flask$^{-1}$) | 19.7 ± 1.6 | 17.6 ± 0.2 | 89.3 |

1989). Following
hybridized for 1 to 2
< SSPE (1× SSPE: 0.
ate, pH 7.7, 1 m*M* E
hardt's solution (Au
DS, and hybridized
ation, the blot was wash
n temperature in 2×
or 15 min at 65°C
olution, and finally
ntaining 0.1% SDS
bridizations were re

tting

eloping kernels and
m cells following the
f (1990). Following
re separated on 0.7%
m*M* Tris-acetate,
piece of Nylon-N
3 to 20 h by capillary
he blot was prehybrid
ontaining 0.5 *M* Na
$ml^{-1}$ salmon sperm
at 65°C for 16 to 20
lenatured *Bt1* cDNA
n blotting. After hybrid
oom temperature for
NaCl, 15 m*M* sodium
1993) plus 0.5% SDS
S, 30 min in 0.1× SSC
°C for 30 min in 0.1×
diography.

mulation in A636 endosperm

ne expression in maize
d cells, we characteri
536 derived endosperm
ycle, the fresh weight of
reased from approxima
spension)$^{-1}$ (Tab. 1).

ltures were subcultured into
t density of approximately

polysaccharides made up about 2.4% of the fresh weight, with starch accounting for 62% of the total polysaccharides (Tab. 1). CCC (100 µ*M*), an inhibitor of gibberellic acid ($GA_3$) biosynthesis (Rademacher 1991), caused an increase in fresh weight (16%), starch (45%), and structural polysaccharides (32%), but α-amylase activity remained essentially unchanged (Tab. 1). These results indicate that A636 endosperm cultures were competent for starch biosynthesis and accumulation.

### Isolation of amyloplast membranes from A636 endosperm cultures

We recently reported that BT1 is a family of abundant polypeptides specific to amyloplast envelope membranes (Cao et al. 1995). Amyloplasts and amyloplast membranes were isolated from A636 suspension-cultured endosperm cells to confirm the presence of BT1 in membranes of in vitro cultured maize endosperm cells. Amyloplasts in these cultures were smaller than those in normally developing maize kernels and in addition to the amyloplasts in the pellet substantial quantities of amyloplasts were also located at the interfaces of the 20/30% and 30/40% Percoll layers during the discontinuous gradient purification. Thus for this study amyloplasts from the two interfaces were combined with those in the pellet. Marker enzyme analyses (Tab. 2) indicate that the combined "purified amyloplast fraction" was free of the cytosol marker, alcohol dehydrogenase, and was only slightly contaminated by mitochondria (1.6% of the total cytochrome *c* oxidase activity), ER (1.4% of the total NADH-cytochrome *c* reductase activity), and microbodies (4.4% of the total catalase activity).

Following removal of starch granules and the stromal fractions, the membrane fractions derived from the purified amyloplasts were yellow in color, a positive marker for amyloplast envelopes from maize endosperm (Cao et al. 1995), potato tubers (Fishwick and Wright 1980), and sycamore suspension-cultured cells (Alban et al. 1988, Harnasut et al. 1988). In addition since the "amyloplast membrane fraction" was isolated from purified amyloplasts it is concluded that this membrane fraction is highly enriched in amyloplast membranes.

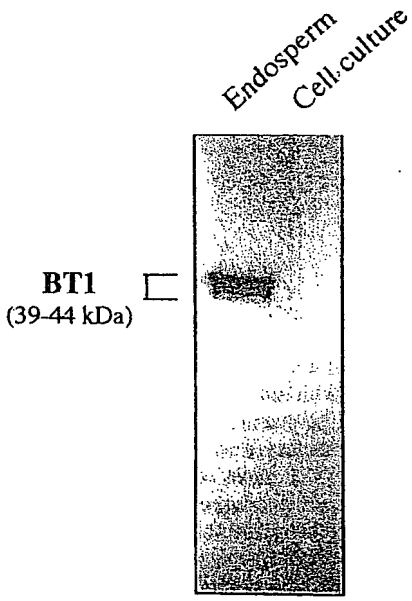



Fig. 1. Western blot analysis of BT1 in amyloplast membranes isolated from the endosperm of 12- to 15-DPP A636 kernels and from suspension-cultured A636 endosperm cells. Each lane was loaded with 4 µg of amyloplast membrane protein.

### BT1 not detected in amyloplast membranes from A636 endosperm cultures

As reported for amyloplast membranes from W64A (Cao et al. 1995), amyloplast membranes isolated from the endosperm of 12- to 15-DPP A636 kernels contained polypeptides (39 to 44 kDa) which reacted with the BT1 antibodies (Fig. 1). To our surprise, amyloplast membranes isolated from suspension-cultured maize endosperm cells lacked detectable levels of BT1 antibody-reacting polypeptides (Fig. 1).

### Comparison of *Bt1* gene expression in A636 endosperm cells growing in vivo and in vitro

Previously we reported that immunoblots (westerns) of polypeptides from amyloplast membranes and total mi-

Tab. 2. Purity of amyloplasts isolated from A636 endosperm cultures. Amyloplasts were isolated from A636 endosperm cultures as described in Materials and methods. Protein and marker enzymes were assayed following filtration through 10-µm pore-sized nylon (homogenate), in the first 100 *g* centrifugation pellet (crude amyloplast), and in the 100 *g* Percoll density gradient fraction (purified amyloplast). All samples were suspended in homogenization buffer and lysed by one freeze and thaw cycle before protein content and marker enzyme activities were determined. Data are the mean plus SE of 2 determinations and are presented as nmol or µmol per minute per gram of fresh weight of endosperm cells as indicated. * Cyanide-insensitive portion.

| % of Control | Preparation | Protein % | Alcohol dehydrogenase (nmol $g^{-1}$ FW $min^{-1}$) | Cyt-c oxidase (nmol $g^{-1}$ FW $min^{-1}$) | NADH-cyt-c reductase* (nmol $g^{-1}$ FW $min^{-1}$) | Catalase (µmol $g^{-1}$ FW $min^{-1}$) |
|---|---|---|---|---|---|---|
| 116.1 | Homogenate | 100 | 163.2 ± 10.0 | 30.8 ± 1.7 | 387.7 ± 108.8 | 24.9 ± 2.7 |
| 137.2 | Crude amyloplast | 18.9 ± 2.0 | 10.5 ± 1.0 | 7.2 ± 1.7 | 125.7 ± 20.0 | 13.2 ± 0.6 |
| 144.8 | Purified amyloplast | 1.3 ± 0.1 | 0 | 0.5 ± 0.1 | 5.3 ± 0.8 | 1.1 ± 0.1 |
| 132.6 | | | | | | |
| 89.3 | Compartment | | Cytosol | Mitochondria | ER | Microbodies |

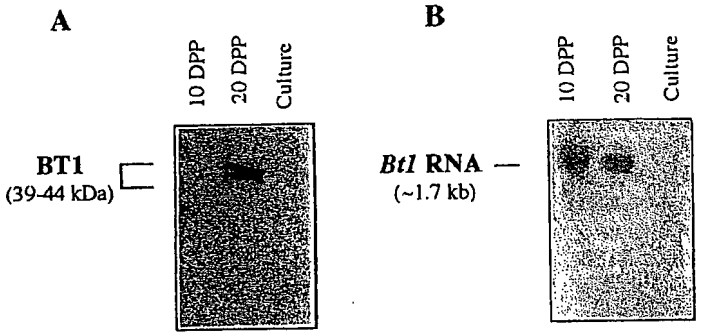



Fig. 2. Western and northern blot analyses of BT1 and *Btl* gene expression, respectively, in extracts from A636 kernels and suspension-cultured A636 endosperm cells. A, Western blot analysis of BT1 in microsomal membranes from A636 kernels and cells from A636 endosperm suspension cultures. Each lane was loaded with 30 µg of microsomal membrane protein. B, Northern blot analysis of *Btl* transcripts in A636 kernels and cells from A636 endosperm suspension cultures. Left and right lanes were loaded with 12 µg of total RNA, mid-dle lane with 4 µg of total RNA.

crosomal membranes have identical banding patterns when reacted with antibodies to BT1 (Cao et al. 1995). Thus to increase the total yield of amyloplast membranes we isolated microsomal membranes from the endosperm of A636 kernels 10 and 20 DPP and from the suspension-cultured endosperm cells, originally initiated by culturing endosperm from 10-DPP A636 kernels. BT1 antibody-reacting polypeptides were present in microsomal membranes from 20-DPP kernels but were not detected in microsomal membranes from 10-DPP A636 kernels or from A636 suspension-cultured endosperm cells (Fig. 2A). The lack of detectable BT1 in microsomal membranes from 10-DPP kernels and A636 endosperm cultures suggests a possibility that cell cultures derived from 10- or 12-DPP kernels retained the same developmental stage as cells in the endosperm of young kernels. To test this possibility, we isolated total RNA from samples identical to those used for the western blots in Fig. 2A. Northern blot analysis indicated that high levels of *Btl* transcripts were detected in both 10- and 20-DPP kernels (Fig. 2B) while only a very weak band was detected in extracts from the cell culture samples. (This very weak band, although not visible in Fig. 2B, may be seen in Fig. 3B.) Southern blotting analysis showed DNA fragments of similar size in developing kernels and in cells from endosperm cultures hybridized with the *Btl* cDNA probe (data not shown), suggesting that the *Btl* gene was present in the endosperm cultures. These results indicate that the *Btl* gene may be differentially expressed in endosperm cells growing in vivo and in vitro.

### *Btl* gene products are not induced by sucrose or CCC

Starch accumulation in suspension-cultured maize endosperm cells is enhanced by the addition of CCC to the medium (Tab. 1) and by supplementation of the culture with additional sucrose (Chu and Shannon 1975). To determine if increased starch synthesis in cultures was related with *Btl* gene expression, we examined the effects of sucrose, CCC, and sucrose plus CCC on *Btl* gene expression in suspension-cultured maize endosperm cells. None of the treatments designed to increase starch synthesis in the maize endosperm cultures affected gene expression (Fig. 3B) or translation (Fig. 3A).

### Comparison of *Btl* gene expression in *waxy* endosperm cells growing in vivo and in vitro

To determine if the reduced *Btl* gene expression was characteristic of the 7-year-old A636 endosperm cultures, we investigated *Btl* gene expression in suspension-cultured *wx* endosperm cultures initiated 3 years before use of the *wx* mutant ge

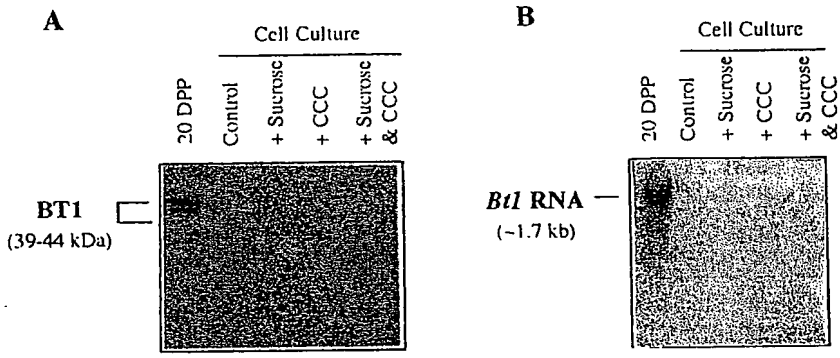
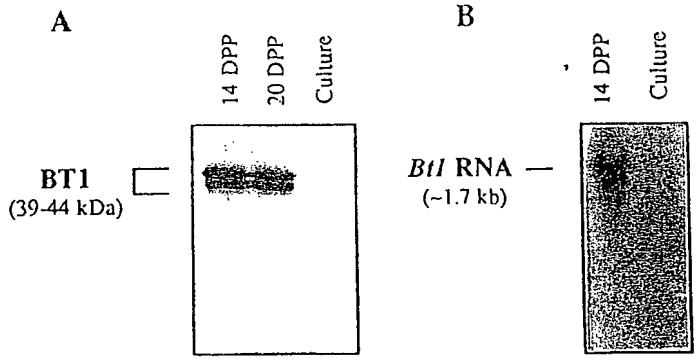



Fig. 3. *Btl* gene expression and translation in A636 suspension-cultured endosperm cells following supplementation with sucrose and the addition of CCC. A, Western blot analysis of BT1 in microsomal membranes from A636 endosperm suspension cultured cells. Polypeptides from microsomal membranes of a 20-DPP kernel are included as a control. Each lane was loaded with 30 µg of microsomal membrane protein. B, Northern blot analysis of *Btl* transcripts in cells from A636 endosperm suspension cultures and 20-DPP kernel (control). Each lane was loaded with 12 µg of total RNA.

## Discussion

BT1, the product of *Btl*, (Cao et al. 1991), appears to be involved in the biosynthesis of starch in maize kernels because in the *btl* mutant genotype, starch is reduced by about 80% (Nelson and Dannenhoffer 1992). In an earlier paper (Cao et al. 1995) we presented evidence supporting the hypothesis that BT1 is a translocator and suggested that it mediates the transfer of ATP, a substrate for starch synthesis, which accounts for about 40% of the ATP used by amyloplast membranes from developing endosperm chloroplasts and other plastids in endosperm, lack detectable BT1 (Cao et al. 1995). If BT1 is a translocator which functions in the transfer of ATP in exchange for ADP, a substrate for amyloplast membranes, then one would be expected to occur in either amyloplasts or chloroplasts. A BT1 homolog not recognized in this study or may be a unique amyloplast membrane protein involved in the transfer of substrates for starch synthesis. ADPGPase is localized in amyloplasts and is generally accepted (Okita et al. 1989) to be in amyloplast membranes but not a translocator. Amyloplasts in synthetic tissues have been shown to use as substrates for




4. Western and northern blot analyses of *Btl* gene expression, respectively, in ...cts from *wx* mutant kernels and suspens... cultured endosperm cells initiated from *wx* ...els. A, Western blot analysis of BT1 in mi... membranes from *wx* kernels and cells ...x endosperm suspension cultures. Each ...was loaded with 30 µg of microsomal ...brane protein. B, Northern blot analysis of ...transcripts in *wx* kernels and cells from *wx* ...sperm suspension cultures. Each lane was ...ed with 12 µg of total RNA.

...-cultured *wx* endosperm cells. These cultures were ...ated 3 years before use, from 12-DPP endosperm tis... of the *wx* mutant genotype in an A636 background. ...levels of BT1 and *Btl* transcripts were detected by ...estern and northern blotting, respectively, in en... from normally developing 14-DPP *wx* kernels, ...no BT1 (Fig. 4A) and only a trace of *Btl* message ...4B) was detected in extracts from the *wx* en... cultures. Genomic Southern blotting showed ...the *Btl* gene was conserved in the *wx* endosperm ...ures (data not shown).

## ...cussion

...the product of *Btl*, a nuclear encoded gene (Sulli... et al. 1991), appears to have an important function ...ine biosynthesis of starch in normally developing ...ize kernels because in its absence, i.e. in kernels of ...*bt1* mutant genotype, starch accumulation is reduced ...about 80% (Nelson and Pan 1995, Tobias et al. 1992). ...In earlier paper (Cao et al. 1995) we reviewed evi... supporting the hypothesis that BT1 is an adenylate ...translocator and suggested that it may function in the ...transfer of ATP, a substrate for AGPase. Although BT1 ...counts for about 40% of the polypeptides in amylo... membranes from maize endosperm, membranes ...from chloroplasts and tissues other than maize en... lack detectable quantities of BT1 polypeptide ...(Cao et al. 1995). If BT1 is an adenylate translocator ...which functions in the transfer of ATP into amyloplasts ...exchange for ADP, a product of starch synthase, then ...amyloplast membranes, regardless of the tissue source, ...would be expected to contain an adenylate translocator. ...thus, either amyloplasts from other tissues may contain ...BT1 homolog not recognized by the BT1 antibodies ...ed in this study or maize endosperm amyloplasts have ...unique amyloplast membrane polypeptide functioning ...the transfer of substrates for starch synthesis.

If AGPase is localized in the amyloplast stroma as ...enerally accepted (Okita 1992, Smith et al. 1995), the ...amyloplast membranes must also contain a carbohy... translocator. Amyloplasts from several nonphoto... synthetic tissues have been shown to take up hexose-Ps ...for use as substrates for starch synthesis (Borchert et al. 1989, Heldt et al. 1991, Hill and Smith 1991, Kosegarten and Mengel 1994, Schott et al. 1995). Amyloplasts isolated from young maize endosperm readily took up and converted ADP-Glc to starch, but they were not able to take up hexose-Ps to be used as substrates for starch synthesis (Liu et al. 1992). Thus the preferred carbohydrate transferred into plastids to provide substrate for starch synthesis varies with the tissue and species. In fact, the specificity of translocators may be modified by environmental or experimental conditions. For example, spinach chloroplasts specifically transfer triose-Ps in exchange for inorganic phosphate (Flugge and Heldt 1991), but it was recently shown that extended incubation of spinach leaves in glucose induced formation of a phosphate translocator capable of transferring Glc-6-P (Quick et al. 1995). In addition, amyloplast membranes from developing potato tubers specifically transferred Glc-6-P, and not Glc-1-P, in exchange for $P_i$ (Schott et al. 1995) while amyloplast membranes from potato tuber cells grown in suspension cultures transferred Glc-1-P rather than Glc-6-P (Kosegarten and Mengel 1994).

Carbohydrate metabolism in suspension-cultured maize endosperm cells differs from that in endosperm cells of intact kernels. For example, the cultured cells produce transient starch; i.e. low quantities of starch accumulate early in the culture cycle and are then metabolized as the medium sugar is depleted (Chu and Shannon 1975). High $\alpha$-amylase activity was observed in the medium and in the suspension-cultured maize endosperm cells (G. Mohabir, K.-C. Liu, C. D. Boyer and J. C. Shannon, unpublished data), but relatively little $\alpha$-amylase activity was detected in the endosperm of developing kernels (Pharis and King 1985). Yu et al. (1991) reported that starvation of rice suspension cultures for only 2–4 h induced $\alpha$-amylase gene expression. The quantity of transient starch in the maize endosperm cultures was increased by supplementation of the cultures with additional sugar (Chu and Shannon 1975) and the addition of CCC (Tab. 1). The results shown here indicate that the *Btl* gene present in suspension-cultured maize endosperm cells was only very weakly expressed and no BT1 product was detectable in membranes from control or sucrose/CCC-supplemented cultures. Thus starch ac-

cumulation in suspension-cultured maize endosperm cells was not dependent on the presence of BT1 in the amyloplast membranes. At present the reasons for differential *Bt1* gene expression in maize endosperm cells growing in vivo and in vitro are unknown. However, the in vitro-grown maize endosperm cells should be useful for studies designed to identify and characterize alternative membrane polypeptides functioning in the transport of substrates for starch synthesis.

*Acknowledgments* – We thank Dr Thomas D. Sullivan (University of Wisconsin, Madison) for the BT1 antibody and the pBt1cDNA1.7 plasmid, Dr Kang-Chien Liu, Christina Richards, Dane K. Fisher, Ming Gao and Kyung-Nam Kim for technical assistance.

## References

Alban, C., Joyard, J. & Douce, R. 1988. Preparation and characterization of envelope membranes from nongreen plastids. – Plant Physiol. 88: 709–717.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. 1993. Current Protocols in Molecular Biology. – Greene Publishing Associates Wiley Interscience, New York, NY. ISBN 0-471-50338-X.

Borchert, S., Grobe, H. & Heldt, H. W. 1989. Specific transport of inorganic phosphate, glucose 6-phosphate, dihydroxyacetone phosphate and 3-phosphoglycerate into amyloplasts from pea roots. – FEBS Lett. 253: 183–186.

Boyer, C. D., Shannon, J. C. & Hardison, R. C. 1989. Physiology, Biochemistry, and Genetics of Nongreen Plastids. – Am. Soc. Plant Physiologists, Rockville, MD. ISBN 0-943088-15-1.

Cao, H., Hu, Z. A., Tan, K. & Jin, Z.-H. 1996. DNA homologues in some species and expression patterns in cucumber organs of genes corresponding to cDNAs from carrot somatic embryogenesis cultures. – J. Plant Physiol. (In press).

– , Sullivan, T. D., Boyer, C. D. & Shannon, J. C. 1995. *Bt1*, a structural gene for the major 39–44 kDa amyloplast membrane polypeptides. – Physiol. Plant. 95: 176–186.

Chu, L.-J. C. & Shannon, J. C. 1975. In vitro cultures of maize endosperm – a model system for studying in vivo starch biosynthesis. – Crop Sci. 15: 814–819.

de Vries, S., Hoge, H. & Bisseling, T. 1988. Isolation of total and polysomal RNA from plant tissues. – *In* Plant Molecular Biology Manual (S. B. Gelvin, R. A. Schilperoort and D. P. S. Verma, eds), pp. 1–13. Kluwer Academic Publishers, Dordrecht. ISBN 0-7923-0236-2.

Doehlert, D. C. & Duke, S. H. 1983. Specific determination of $\alpha$-amylase activity in crude plant extracts containing $\beta$-amylase. – Plant Physiol. 71: 229–234.

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. 1956. Colorimetric method for determination of sugars and related substances. – Anal. Chem. 28: 350.

Felker, C. F. 1987. Ultrastructure of maize endosperm suspension cultures. – Am. J. Bot. 74: 1912–1920.

– & Goodwin, J. C. 1988. Sugar uptake by maize endosperm suspension cultures. – Plant Physiol. 88: 1235–1239.

– , Miernyk, J. A. & Crawford, C. G. 1989. Characterization of maize endosperm-derived suspension cells through the culture cycle and enhancement of tissue friability. – Plant Cell Tissue Organ Cult. 18: 153–165.

Fishwick, M. T. & Wright, A. J. 1980. Isolation and characterization of amyloplast envelope membranes from *Solanum tuberosum*. – Phytochemistry 19: 55–59.

Flugge, U.-I. & Heldt, H. W. 1991. Metabolite translocators of the chloroplast envelope. – Annu. Rev. Plant Physiol. Plant Mol. Biol. 42: 129–144.

Fourney, R. M., Miyakoshi, J., Dy, R. S. III & Paterson, M[illegible] 1990. Northern blotting: efficient RNA staining and tran[illegible] – Focus 10: 5–7.

Hannah, L. C., Giroux, M. & Boyer, C. D. 1993. Biotechnolo[illegible] cal modification of carbohydrate for sweet maize impro[illegible] ment. – Sci. Hortic. 55: 177–197.

Harinasut, P., Akazawa, T., Zhou, F.-G. & Takabe, T. 19[illegible] Polypeptide compositions of stroma, internal membran[illegible] and inner and outer envelope membranes of amylo[illegible] from suspension-cultured cells of sycamore (*Acer pseu[illegible] platanus* L.). – Plant Cell Physiol. 29: 1315–1321.

Heldt, H. W., Flugge, U.-I. & Borchert, S. 1991. Diversity [illegible] specificity and function of phosphate translocators in [illegible] ous plastids. – Plant Physiol. 95: 341–343.

Hill, L. M. & Smith, A. M. 1991. Evidence that glucose 6-pho[illegible] phate is imported as the substrate for starch synthesis by [illegible] plastids of developing pea embryos. – Planta 185: 91–96.

Hodge, J. E. & Hofreiter, B. T. 1962. Determination of reduc[illegible] sugars and carbohydrates. – *In* Methods in Carbohyd[illegible] Chemistry (R. L. Whistler and M. L. Wolfrom, eds), Vol. [illegible] pp. 380–394. Academic Press, New York, NY. ISBN 0-[illegible] 746206-6.

Junghans, H. & Metzlaff, M. 1990. A simple and rapid meth[illegible] for the preparation of total plant DNA. – Biotechniques [illegible] 176.

Keegstra, K., Olsen, L. J. & Theg, S. M. 1989. Chloroplastic [illegible] cursors and their transport across the envelope membran[illegible] Annu. Rev. Plant Physiol. Plant Mol. Biol. 40: 471–501.

Kosegarten, H. & Mengel, K. 1994. Evidence for a glucos[illegible] phosphate translocator in storage tissue amyloplasts o[illegible] tato (*Solanum tuberosum*) suspension-cultured cells. – [illegible] iol. Plant. 91: 111–120.

Lee, S.-Y. & Rasheed, S. 1990. A simple procedure for ma[illegible] mum yield of high-quality plasmid DNA. – Biotechniques [illegible] 676–679.

Li, H.-M., Sullivan, T. D. & Keegstra, K. 1992. Information [illegible] , targeting to the chloroplastic inner envelope membran[illegible] contained in the mature region of the maize *Bt1*-encoded [illegible] protein. – J. Biol. Chem. 267: 18999–19004.

Liu, K.-C., Boyer, C. D. & Shannon, J. C. 1992. Carbohydr[illegible] transfer into isolated maize (*Zea mays* L.) amyloplasts (A[illegible] stract No. 234). – Plant Physiol. 99: S–39.

Mangelsdorf, P. C. 1926. The genetics and morphology of [illegible] endosperm characters in maize. – Conn. Agric. Exp. [illegible] Bull. 279: 509–614.

Nelson, O. & Pan, D. 1995. Starch synthesis in maize [illegible] dosperms. – Annu. Rev. Plant Physiol. Plant Mol. Biol. [illegible] 475–496.

Okita, T. W. 1992. Is there an alternative pathway for starch [illegible] thesis? – Plant Physiol. 100: 560–564.

Pharis, R. P. & King, R. W. 1985. Gibberellins and reproduct[illegible] development in seed plants. – Annu. Rev. Plant Physiol. [illegible] 517–568.

Preiss, J. 1991. Biology and molecular biology of starch [illegible] thesis and its regulation. – *In* Oxford Survey of Plant Mo[illegible] lecular and Cellular Biology, Vol. 7 (B. J. Miflin, ed.), [illegible] 59–114. Oxford University Press, Oxford. ISBN 0-[illegible] 857750-8.

Quick, W. P., Scheibe, R. & Neuhaus, H. E. 1995. Induction [illegible] hexose-phosphate translocator activity in spinach chlor[illegible] plasts. – Plant Physiol. 109: 113–121.

Rademacher, W. 1991. Inhibitors of gibberellin biosynthesis: [illegible] plications in agriculture and horticulture. – *In* Gibberellins [illegible] (N. Takahashi, B. O. Phinney and J. MacMillan, eds), [illegible] 296–310. Springer-Verlag, New York, NY. ISBN 0-[illegible] 97259-5.

Sambrook, J., Fritsch, E. F. & Maniatis, T. 1989. Molec[illegible] Cloning: A Laboratory Manual, 2 Ed. – Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. pp 18.47–18.74. ISBN 0-87969-309-6.

Saravitz, C. H. & Boyer, C. D. 1987. Starch characteristi[illegible] cultures of normal and mutant maize endosperm. – Theor. Appl. Genet. 73: 489–495.

[illegible] nell, D. J., Kessler, F. & [illegible] ponents of the chloropla[illegible] ence 266: 1007–1012.

[illegible] ott, K., Borchert, S., Mul[illegible] Transport of inorganic [illegible] phosphates across the e[illegible] amyloplasts. – Planta 19[illegible]

[illegible] hannon, J. C. 1982. Maize [illegible] Biological Research (W[illegible] Plant Molecular Biolog[illegible] ISBN 0-9608758-0-8.

[illegible] Echeverria, E. & Boye[illegible] plasts from developing e[illegible] Methods Enzymol. 148:

[illegible] Pien, F.-M. & Liu, K.-C[illegible] sugars in developing ma[illegible] thesis of ADP-glucose [illegible] 835–843.

[illegible] Smith A. M., Denyer K. & M[illegible] amount and structure o[illegible] Physiol. 107: 673–677.

[illegible] Sullivan, T. D. & Kaneko, Y[illegible] codes amyloplast mem[illegible] 477–484.

Edited by T. M. Murphy

, R. S. III & Paterson, M.
nt RNA staining and tra

r, C. D. 1993. Biotechnol
ite for sweet maize imp
17.
, F.-G. & Takabe, T.
stroma, internal membr
membranes of amylo
of sycamore (*Acer*
iol. 29: 1315–1321.
rchert, S. 1991. Diversity
osphate translocators
5: 341–343.
Evidence that glucose 6-ph
ite for starch synthesis by
ryos. – Planta 185: 91–96.
2. Determination of red
In Methods in Carbohyd
M. L. Wolfrom, eds). Vol
New York, NY. ISBN 0

. A simple and rapid meth
nt DNA. – Biotechniques

S. M. 1989. Chloroplastic
ss the envelope membra
Mol. Biol. 40: 471–501.
4. Evidence for a gluco
ge tissue amyloplasts of
ension-cultured cells.

simple procedure for
mid DNA. – Biotechniq

stra, K. 1992. Informatio
inner envelope membra
n of the maize *Bt1*-encod
18999–19004.
on, J. C. 1992. Carbohyd
*ea mays* L.) amyloplasts
l. 99: S–39.
tics and morphology of
e. – Conn. Agric. Exp.

arch synthesis in maize
Physiol. Plant Mol. Biol.

native pathway for starch
50–564.
Gibberellins and reproduct
Annu. Rev. Plant Physiol.

ecular biology of starch
Oxford Survey of Plant
Vol. 7 (B. J. Miflin, ed.)
Press, Oxford. ISBN 0

aus, H. E. 1995. Induction
r activity in spinach chlo
3–121.
f gibberellin biosynthesis
orticulture. – *In* Gibberell
and J. MacMillan, eds).
ew York, NY. ISBN 0

Maniatis, T. 1989. Molecu
1, 2 Ed. – Cold Spring Har
or, NY. pp 18.47–18.74. ISB

987. Starch characteristics
it maize endosperm. – The

Schnell, D. J., Kessler, F. & Blobel, G. 1994. Isolation of components of the chloroplast protein import machinery. – Science 266: 1007–1012.
Schott, K., Borchert, S., Muller-Rober, B. & Heldt, H. W. 1995. Transport of inorganic phosphate and C3- and C6-sugar phosphates across the envelope membranes of potato tuber amyloplasts. – Planta 196: 647–652.
Shannon, J. C. 1982. Maize endosperm cultures. – *In* Maize for Biological Research (W. F. Sheridan, ed.), pp. 397–400. Plant Molecular Biology Association, Charlottesville, VA. ISBN 0-9608758-0-8.
Echeverria, E. & Boyer, C. D. 1987. Isolation of amyloplasts from developing endosperm of maize (*Zea mays* L.). – Methods Enzymol. 148: 226–234.
Pien, F.-M. & Liu, K.-C. 1996. Nucleotides and nucleotide sugars in developing maize (*Zea mays* L.) endosperms: Synthesis of ADP-glucose in *brittle*-1. – Plant Physiol. 110: 835–843.
Smith A. M., Denyer K. & Martin C. R. 1995. What controls the amount and structure of starch in storage organs. – Plant Physiol. 107: 673–677.
Sullivan, T. D. & Kaneko, Y. 1995. The maize *brittle*1 gene encodes amyloplast membrane polypeptides. – Planta 196: 477–484.
– , Strelow, L. I., Illingworth, C. A., Phillips, R. L. & Nelson, O. E. Jr. 1991. Analysis of maize *Brittle*-1 alleles and a defective suppressor-mutator-induced mutable allele. – Plant Cell 3: 1337–1348.
Takahashi, R., Valeika, S. R. & Glass, K. W. 1992. A simple method of plasmid transformation of *E. coli* by rapid freezing. – Biotechniques 13: 711–715.
Tobias, R. B., Boyer, C. D. & Shannon, J. C. 1992. Alteration in carbohydrate intermediates in the endosperm of starch-deficient maize (*Zea mays* L.) genotypes. – Plant Physiol. 99: 146–152.
Villand, P. & Kleczkowski, L. A. 1994. Is there an alternative pathway for starch biosynthesis in cereal seeds? – Z. Naturforsch 49c: 215–219.
Wentz, J. B. 1926. Heritable characters in maize. XXVI-concave. – J. Hered. 17: 327–329.
Yu, S.-M., Kuo, Y.-H., Sheu, G., Sheu, Y.-J. & Liu, L.-F. 1991. Metabolic derepression of α-amylase gene expression in suspension-cultured cells of rice. – J. Biol. Chem. 266: 21131–21137.

Edited by T. M. Murphy

# Discovering Empirically Conserved Amino Acid Substitution Groups in Databases of Protein Families

**Thomas D. Wu and Douglas L. Brutlag**

Department of Medicine, Section on Medical Informatics,
and Department of Biochemistry
Stanford University Medical Center
Stanford, California 94305
twu@camis.stanford.edu, brutlag@cmgm.stanford.edu

**Abstract**

This paper introduces a method for identifying empirically conserved amino acid substitution groups. In contrast with existing approaches that view amino acid substitution as a pairwise phenomenon, the method presented here identifies conserved groups of amino acids using a data structure called a conditional distribution matrix. The conditional distribution matrix extends the concept of a pairwise substitution matrix by changing the context of substitution from a single amino acid to a group of amino acids. The matrix tabulates information from a database of protein families that contains numerous aligned positions. Each row in the matrix contains the distribution of amino acids in those aligned positions that contain a given conditioning group of amino acids. The method converts a database of protein families into a conditional distribution matrix and then examines each possible substitution group for evidence of conservation. The algorithm is applied to the BLOCKS and HSSP databases. Twenty amino acid substitution groups are found to be conserved empirically in both databases. These groups provide insight into biochemical properties that are conserved in protein evolution.

## Introduction

Protein sequences often exhibit variability in their amino acid patterns, a phenomenon that can be characterized as substitutions of amino acids for one another. Amino acid substitution has been viewed largely as a pairwise phenomenon. Typically, this phenomenon is represented by the frequency that one amino acid replaces another one; the entire set of replacement or substitution frequencies is organized into a symmetric 20 x 20 substitution matrix containing 210 distinct pairwise frequencies. Substitution matrices have been studied in depth (Altschul 1991; Gonnet, Cohen, & Benner 1992; Jones, Taylor, & Thornton 1992; Vogt, Etzold, & Argos 1995), and various matrices have been proposed, including the well-known accepted point mutation (PAM) matrix of Dayhoff and colleagues (1978). In many cases, such substitution matrices have proven quite useful for comparing, aligning, and exploring relationships between pairs of protein sequences.

However, for groupwise or consensus relationships, statistics and methods based on pairwise comparisons are often inadequate. The shift from pairwise comparisons to groupwise analyses is often challenging and non-trivial, as can be seen from the difficulties in trying to align multiple sequences (Barton 1990). Sometimes such problems can be approached using pairwise methods, but often new methods are needed. In computational biology, much attention has focused recently on groupwise or consensus analyses, such as the classification of families and superfamilies of protein sequences and structures, and the compilation of protein family databases, such as PROSITE (Bairoch 1991), BLOCKS (Henikoff & Henikoff 1991), and HSSP (Sander & Schneider 1991)

In light of these advances in computational biology, we present in this paper an empirical analysis of amino acid substitution using a group perspective. We introduce a novel method for identifying groups of amino acids that substitute for one another with high frequency. Our method identifies these substitution groups empirically from a collection of multiple sequence alignments. Although some researchers have also used multiple sequence alignments to study amino acid substitution (Henikoff & Henikoff 1992), their goal has been to derive new substitution matrices, whereas our goal is to identify substitution groups.

Various classifications of amino acids into meaningful and useful groups have been proposed in other studies, as summarized in Table 1. However, previous methods for identifying substitution groups have either analyzed pairwise data or used theoretical principles rather than empirical data. Some researchers have used pairwise substitution matrices to infer substitution groups (Dayhoff, Schwartz, & Orcutt 1978; Miyata, Miyazawa, & Yasunaga 1979). A major problem with such an approach is that substitutability is not necessarily transitive. That is, even if amino acids A and B substitute for each other in some contexts and amino acids B and C substitute for each other in other contexts, we cannot automatically conclude that amino acids A and C substitute for each other. Another problem with pairwise analyses is that they are limited in their ability to distinguish different biochemical contexts

| Reference | Substitution groups |
|---|---|
| Dayhoff et al. 1978 | C, FWY, HKR, DENQ, ILMV, AGPST |
| Miyata et al. 1979 | C, FWY, HKR, DENQ, ILMV, AGPST |
| Jimenez-Montano & Zamora-Cortina 1981 | AG, DE, KR, NQ, ST, FWY, ILMV, CFILMVWY, ADEGHKNPQRST |
| Taylor 1986 | Approximately 70 union and intersection combinations of HKR, ILV, ACGS, HFWY, DEHKR, ACDGNPSTV, CDEHKNQRSTWY, ACFGHIKLMTVWY |
| Smith & Smith 1990 | P, AG, DE, NQ, ST, FWY, HKR, ILV, CFILMVWY, DEHKNQRST |
| Mocz 1995 | AEHMQY, FIKLVW, CDGNPRST |
| Naor et al. 1996 | DN, GP, DGNP, EKQR, FILV, DEKNQR, ACFILMVWY, DEGHKNPQRST |
| Klingler 1996 | H, K, N, AP, CF, DE, GS, KR, AGS, ILV, NQR, QTY, HMTWY, CFILMVW |

**Table 1** Amino acid substitution groups identified in previous studies. Substitution groups are arranged in order of increasing size. Singleton groups, consisting of a single amino acid, are listed when the study indicated that the amino acid was relatively unlikely to substitute for other amino acids.

for substitution. The concept of amino acid substitution inherently requires a context; without a context, we merely have a marginal distribution of amino acid frequencies. Therefore, substitutability essentially consists of rules of the form, "When context X is present, amino acid A substitutes with frequency *f*." With pairwise data, the context X can be specified only as a single amino acid. In contrast, in this paper, we consider more expressive and specific contexts that contain groups of amino acids rather than a single amino acid.

Other researchers have proposed substitution groups on theoretical rather than empirical grounds (Jimenez-Montano & Zamora-Cortina 1981; Kidera et al. 1985; Taylor 1986; Smith RF and Smith TF 1990; Mocz 1995). These theoretical analyses use measurements of various amino acid properties, such as volume, charge, and hydrophobicity, and then propose substitution groups that should be conserved. Unfortunately, theoretical models may not necessarily correspond to the patterns of conservation observed empirically. Moreover, amino acid properties often depend on particular biochemical environments found in protein structures, so properties of amino acid in isolation may not reflect the complexities of amino acid substitutions in particular contexts.

In our approach, we analyze each possible substitution group on its own merits. Hence, substitution groups may overlap or subsume one another. Our approach differs from that of some other researchers, who require that substitution groups do not overlap. For example, Mocz (1995) uses clustering techniques to identify three mutually exclusive clusters of amino acids. The mutual exclusion requirement means that each amino acid can belong only to a single substitution group. Another restriction sometimes placed is that the groups must be organized into a strict hierarchy (Smith RF & Smith TF 1990) or Venn diagram (Taylor 1986). We believe that such requirements are unnecessarily restrictive. Each amino acid has several properties and can serve different functions, depending on the biochemical context. In some contexts, the size of an amino acid may be critical; in others, its charge may be the conserved property. These different contexts will not necessarily fit into a mutually exclusive, strictly hierarchical, or set-theoretical scheme. Therefore, in our approach, we analyze each substitution group separately for empirical evidence of conservation.

In order to analyze each possible substitution group independently, we require more information than is stored in a pairwise substitution matrix. We therefore extend the idea of a substitution matrix to a larger structure called a conditional distribution matrix. This matrix provides the foundation for a consensus-based analysis of amino acid substitution. In the rest of this paper, we discuss two databases of protein families, BLOCKS and HSSP, that make a consensus-based analysis possible. We then present the concept of the conditional distribution matrix and the criteria we use to identify empirically conserved substitution groups. We then present independent analyses of substitution groups conserved empirically in the BLOCKS and HSSP databases. We find that twenty substitution groups are conserved in both databases, and we propose biochemical characteristics underlying those groups. Finally, we discuss various features of our approach and suggest how our results may be used in further work in computational biology.

## Methods

### Data

Our method requires a source of aligned positions; this data is readily available from databases of protein families or multiple sequence alignments. Two of largest and most widely used protein family databases are the BLOCKS and HSSP databases. Although these databases have distinct characteristics, they can still be viewed as collections of aligned positions.

The BLOCKS database (Henikoff & Henikoff 1991) contains short, highly conserved regions of protein families, represented by ungapped multiple alignments called blocks. Blocks are generated from a set of related

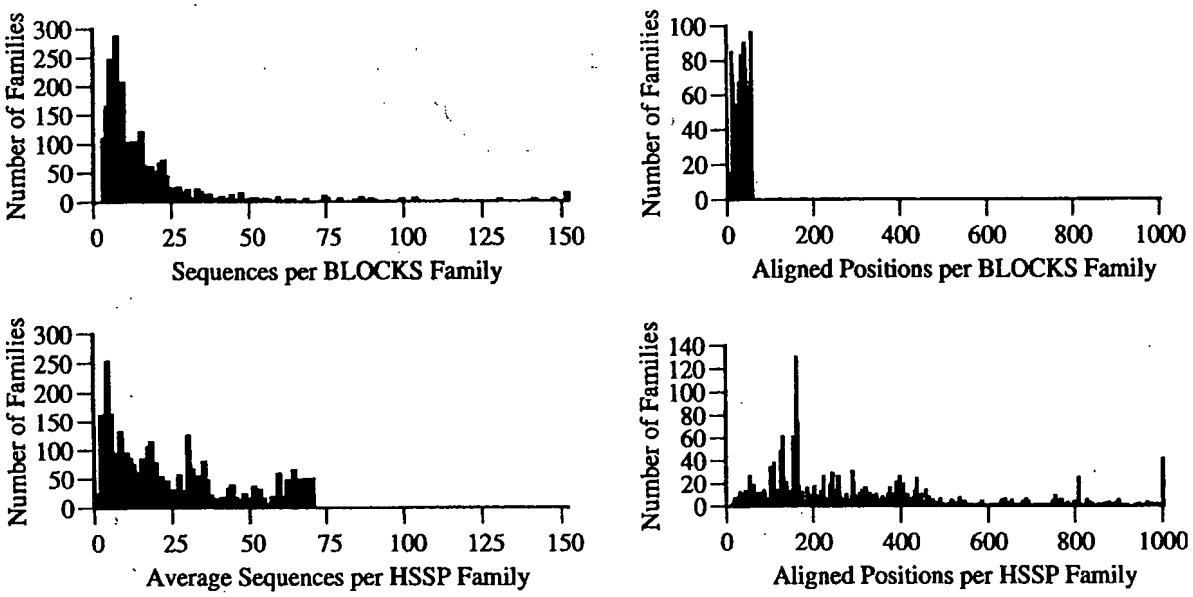



**Figure 1** Comparison of BLOCKS and HSSP databases. The histograms show the number of protein families of different sizes in the two databases, where the size of a family may be measured by the number of sequences or aligned positions that it contains. The horizontal scales are selected to be the same dimension to facilitate comparison. The BLOCKS database actually contains some protein families with as many as 507 sequences, and the HSSP database contains some protein families with as many as 1983 aligned positions.

protein sequences. Conserved regions are then found within these sequences using a motif finding program (Smith HO, Annau, & Chandrasegaran 1990), and the edges of these regions are extended until a similarity score falls below some threshold. The similarity score for local alignment and extension is based on the BLOSUM 62 substitution matrix. Finally, a highly scoring set of blocks is selected from all possible conserved regions using an optimal path algorithm. In our study, we used version 8.0 of the BLOCKS database (9 August 1994), which contains 2884 blocks constructed from 770 protein groups in PROSITE version 12.0.

The HSSP (Homology-derived Secondary Structure of Proteins) database (Sander & Schneider 1991) combines structural data from the PDB (Protein Data Bank) database and sequence data from the SWISSPROT database. Each HSSP family corresponds to a PDB structure, and contains all SWISSPROT sequences that are homologous above a certain length-dependent threshold, using the Smith-Waterman alignment algorithm and a substitution matrix by McLachlan (1971). We used the version of HSSP dated 16 November 1995, which contains 3569 protein families.

The BLOCKS and HSSP databases are constructed in quite different ways for different purposes. The BLOCKS database aims to describe sequence homology, whereas the HSSP database aims to describe structural homology, inferred from sequence homology. The BLOCKS database contains families based on optimally scoring multiple sequence alignments of locally conserved regions, whereas HSSP contains families based on pairwise sequence comparisons over global regions. The BLOCKS database does not allow gaps in its alignments, whereas HSSP does. The two databases use different substitution matrices for computing alignments.

These differences produce different types of multiple sequence alignments, as shown in Figure 1. A protein family can be characterized by the number of sequences and aligned positions it contains. In the HSSP database, different sequences may contribute to each aligned position, so the number of sequences should be averaged over all positions. As Figure 1 shows, the number of sequences per family in BLOCKS varies widely, ranging from 2 to 507 (mean = 15.6), whereas in HSSP, it is relatively narrow, ranging from 1 to 70 (mean = 24.9). Conversely, the number of positions per family in BLOCKS ranges only from 4 to 55 (mean = 32.8), whereas in HSSP, it ranges from 12 to 1983 (mean = 266.9). These histograms reflect the stricter requirement that BLOCKS places on each aligned position, requiring it to be conserved across all sequences. On the other hand, HSSP may include an aligned position that is conserved for some but not all homologous sequences.

## Conditional Distribution Matrix

Our algorithm consists of two steps. First, we convert a database of aligned positions, such as those in BLOCKS and HSSP, into a large data structure called a **conditional distribution matrix** (CDM). Then, we look for statistically significant groups of amino acids within this matrix. In this section, we describe the CDM.

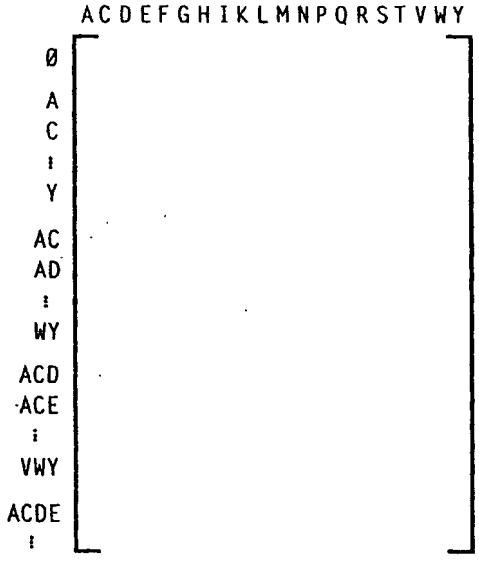



**Figure 2** Conditional distribution matrix. Each row corresponds to a conditioning group, which identifies a subset of the aligned positions in the database. The entries in each row contain the distribution of amino acids in those aligned positions.

The conditional distribution matrix can be thought of as an extension of the 20 x 20 pairwise substitution matrix. The pairwise substitution matrix essentially contains different contexts for amino acid substitution, where each context is a single amino acid. For each context $a$, the pairwise substitution matrix contains the distribution of amino acids that substitute for $a$. Hence, entries in the pairwise substitution matrix are substitution frequencies $f(a'|a)$ that indicate the likelihood that amino acid $a'$ substitutes in the context of amino acid $a$. If we think of the context as being on the vertical axis, then each row contains the distribution for a different context.

If we extend the idea of context from a single amino acid to a set of amino acids, we obtain the conditional distribution matrix (Figure 2). The contextual set of amino acids is called a **conditioning group**. Each conditioning group $A$ corresponds to a subset of the aligned positions in the database. To understand the correspondence, suppose that we represent an aligned position as a set $P$ of amino acids. The set $P$ is the union of all amino acids in the aligned position. Then, we say that an aligned position $P$ **satisfies** conditioning group $A$ if $A \subseteq P$. In other words, the aligned position must contain at least one instance of every amino acid in the conditioning group. The aligned position may, of course, contain other amino acids as well, and in fact, it is the distribution of these other amino acids that we are interested in. The conditioning group $A$ provides a context for substitution, whereas the remaining amino acids in $(P-A)$ substitute in that context.

Note that a given aligned position may satisfy several conditioning groups. For instance, consider an aligned position that contains 5 occurrences of valine, 4 occurrences of leucine, and 3 occurrences of isoleucine. Then, the aligned position satisfies the conditioning groups I L V, I L, I V, L V, I, L, V, and the null set Ø. This multiple correspondence is appropriate because it is difficult to know *a priori* the biochemical context or functionality of an aligned position and hence its substitution pattern. For instance, in this example, the underlying context might be I L V, with no substitutions Or the underlying context might be L V, with isoleucine substituting at a high frequency. Or perhaps the context might simply be V, with leucine and isoleucine both substituting at a high frequency.

Although we may not be able to infer substitution patterns from a single aligned position, sampling over several aligned positions may provide statistically meaningful answers. The conditional distribution matrix accumulates the substitution patterns for various possible contexts over numerous aligned patterns. This matrix, which is of size $2^{20}$ x 20, has a row for each possible conditioning group $A$. Each row contains the distribution of substituting amino acids in the context of a certain conditioning group. Each entry in the CDM contains a conditional count $c(a|A)$, which equals the total number of occurrences of amino acid $a$ over all aligned positions that satisfy conditioning group $A$. We may think of the counting process conceptually as finding all aligned positions that satisfy conditioning group $A$, and then tabulating all amino acids in those positions. (In practice, however, the CDM is not constructed row by row, but by processing each aligned position in a database sequentially. For each aligned position $P$, we add the counts of amino acids in $P$ to all rows $A$, such that $A \subseteq P$.)

Since the number and sizes of aligned positions satisfying each conditioning group varies, we normalize the conditional counts to obtain a conditional frequency:

$$f(a|A) = \frac{c(a|A)}{\sum_{a' \notin A} c(a'|A)}, \quad \text{where } a \notin A.$$

Note that the normalizing value in the denominator excludes amino acids in the conditioning group. This avoids the problem of circularity, whereby the count of an amino acid in the group is elevated simply because the group selects for it. Rather, we are interested in the distribution of amino acids outside the conditioning group.

In order to evaluate these conditional frequencies, we require an expected value for comparison. The expected frequency comes from the marginal distribution of amino acids, that is, the distribution across all aligned positions in the database. In fact, if the null set is a considered a conditioning group, the marginal counts $c(a)$ will be stored in the CDM. The expected conditional frequency derives from the marginal counts as follows:

$$\mu(a|A) = \frac{c(a)}{\sum_{a' \notin A} c(a')}, \quad \text{where } a \notin A.$$

In addition to first-order statistical characteristics, such as the observed and expected frequencies, we may also compute second-order statistical characteristics, such as the standard error of the proportion $\mu$:

$$\hat{\sigma}(a|A) = \sqrt{\frac{\mu(a|A)[1-\mu(a|A)]}{\sum_{a' \in A} c(a'|A)}}$$

Given these first- and second-order quantities, we can evaluate an observed conditional frequency $f$ using the relative deviate or Z-score. The Z-score indicates the number of standard errors $\hat{\sigma}(a|A)$ that an observed frequency $f(a|A)$ differs from the expected frequency $\mu(a|A)$:

$$Z(a|A) = \frac{f(a|A) - \mu(a|A)}{\hat{\sigma}(a|A)}$$

The Z-score indicates whether an amino acid is over- or under-represented in the context of a given conditioning group. If the Z-score is positive, the amino acid is over-represented in that context; and if negative, it is under-represented. We may imagine that each conditioning group induces a frequency distribution on the other amino acids. Amino acids with positive Z-scores are positively induced, whereas those with negative Z-scores are negatively induced. For the purpose of definition, we use three standard errors as a threshold: A Z-score greater than 3 reflects positive induction; a Z-score less than –3, negative induction; and between 3 and –3, a neutral effect.

## Criteria for Empirical Conservation

The conditional distribution matrix can be analyzed for evidence that a given substitution group is conserved. In particular, the Z-scores provide a basis for identifying substitution groups that are empirically conserved. Intuitively, we consider a substitution group $A$ to be conserved empirically if amino acids within the group substitute for one another significantly more frequently than amino acids outside the group substitute for amino acids in $A$.

More formally, we consider a group conserved empirically if it is both compact and isolated. Compactness means that all amino acids in the group substitute for one another frequently, and isolation means that amino acids outside the group do not substitute for those within the group as frequently. We measure substitutability within the group as the Z-score of each amino acid conditioned on other members of the group, or $Z(a|A-\{a\})$. The overall compactness, or **compactness score**, is the minimum of these scores:

$$C(A) = \min_{a \in A} Z(a|A-\{a\})$$

We measure substitutability of amino acids outside the group for those within the group as the Z-score of each amino acid conditioned on the group, or $Z(a|A)$. Because a high score indicates that an amino acid outside the group should belong to the group, we define the **interference score** to be the maximum of these scores:

$$I(A) = \max_{a \notin A} Z(a|A)$$

Finally, we quantify the conservation of a substitution group by the difference between its compactness and interference scores. We call this the **separation score**:

$$S(A) = C(A) - I(A)$$

When a substitution group has a statistically significant separation score, we say that the substitution group is **conserved empirically**. We set the threshold for significance at three standard errors, which is equivalent to a significance level of 0.01. We examined each possible substitution group for a separation score greater than three standard errors.

Because we test a large number of substitution groups independently, one may ask whether the number of tests itself will yield a large number of significant results. Surprisingly, the answer is no. Let the size $N$ be defined as the cardinality of substitution group A, that is, the number of amino acids in $A$. Consider all possible substitution groups of size N; there are "20 choose N" or $20!/[N!\,(20-N)!]$ such groups to be tested. If Z-scores are distributed randomly, then a group has a positive separation score whenever the Z-scores of the $N$ amino acids in the group are all greater than the Z-scores of the $(20-N)$ outside the group. Hence, the probability of achieving a positive separation score for a group of size N is the permutation of N multiplied by the permutation of $(20-N)$, divided by all permutations of the 20 amino acids. This is simply the reciprocal of the number of groups of size N. Therefore, among all substitution groups of size $N$, we expect to see one group with a positive separation score by random chance. Hence, we need not make a provision, such as a Bonferroni correction, for the large number of tests.

## Examples

To gain a better understanding of our method, we look at some examples. Consider the substitution group ILV; our analysis of this group using the BLOCKS database is shown in Table 2(a). The Z-scores for isoleucine, valine, and leucine are all show high rates of substitution for one another, significantly higher than any other amino acid. The closest amino acid that interferes with this group is methionine. Although methionine is positively induced by ILV, there is a clear separation of 102.4 standard errors between the substitution frequencies of ILV and M, which is highly significant. Hence, ILV is conserved empirically in the BLOCKS database. Note that ILV positively induces the hydrophobic amino acids M, F, T, A, and Y; negatively induces Q, C, S, P, W, N, R, E, D, and G; and has a neutral effect on H. The group also positively induces K slightly, which makes biochemical sense because the long side chain of lysine gives it a partially hydrophobic character.

In contrast with ILV, most substitution groups were not conserved empirically. Table 2(b) shows the analysis for the group GIM, which scored the lowest among all substitution groups of size 3 in the BLOCKS database. GIM scores poorly because it is not compact. Glycine

(a) Substitution group I L V (5328 positions) Separation score: 102.4

| I | V | L | M | F | T | A | Y | K | H | Q | C | S | P | W | N | R | E | D | G |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 219.7 | 189.8 | 188.7 | 86.3 | 61.9 | 40.0 | 19.6 | 7.4 | 6.7 | -0.9 | -6.1 | -7.0 | -9.6 | -15.4 | -16.3 | -16.6 | -21.3 | -21.9 | -29.2 | -51.0 |

(b) Substitution group G I M (606 positions) Separation score: -119.4

| M | I | G | L | V | F | A | K | H | S | Q | C | W | Y | T | N | E | R | P | D |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 7.7 | 7.2 | -61.9 | 57.5 | 32.9 | 19.9 | 7.2 | 0.9 | 0.3 | -5.7 | -6.5 | -6.9 | -11.7 | -11.8 | -13.3 | -13.3 | -13.7 | -20.6 | -25.5 | -26.2 |

(c) Substitution group F I V (2080 positions) Separation score: -108.8

| I | V | F | L | M | Y | T | H | A | K | W | C | S | Q | N | D | E | P | R | G |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 96.4 | 73.3 | 31.3 | 140.1 | 33.4 | 24.4 | 10.8 | 0.9 | -2.9 | -4.0 | -6.2 | -8.8 | -14.3 | -16.4 | -17.3 | -25.6 | -29.2 | -31.4 | -33.2 | -47.4 |

**Table 2** Analyses of substitution groups I L V, G I M, and F I V. The amino acids in each group are separated from those outside the group by a double bar, and then sorted by Z-score. For each substitution group $A$ and amino acid $a$, the values to the left of the double bar are $Z(a|A-\{a\})$, and the values to the right of the double bar are $Z(a|A)$. A single vertical bar separates amino acids that are positively, neutrally, and negatively induced by the substitution group.

substitutes only rarely in those positions with both isoleucine and methionine, as shown by its under-representation of 61.9 standard errors. Hence, G I M is not conserved empirically in the BLOCKS database, for reasons of non-compactness.

Another substitution group, F I V, shown in Table 2(c), failed our criterion because it was not isolated. Although its compactness score is relatively high, meaning that the three amino acids each substitute for one another frequently, its interference score is even higher, because leucine substitutes in this context even more frequently. Therefore, the substitution group F I V is also not conserved empirically, for reasons of non-isolation.

## Results

Our analysis of the BLOCKS database yielded 30 substitution groups that are conserved empirically, and our analysis of the HSSP database yielded 51 substitution groups. These substitution groups are listed in Tables 3 and 4, respectively. Twenty substitution groups are conserved empirically in both databases. We feel that the validation of these substitution groups by both databases provides strong evidence that they are indeed conserved in nature. We therefore consider further the biochemical characteristics of these substitution groups.

Of the 190 possible amino acid groups of size 2, nine are conserved empirically in both databases. These amino acid pairs are not evident immediately from substitution matrices. For example, the empirically conserved substitution groups have the following BLOSUM 62 scores: F Y (score of 3), I V (3), S T (1), A S (1), D E (2), K R (2), D N (1), E Q (2), and H Y (2). Conversely, the BLOSUM 62 matrix contains several positively scoring amino acid pairs that were not conserved empirically in our study: L M, I L, and W Y (all with scores of 2), and N S, N H, R Q, E K, K Q, I M, M V, L V, and F W (all with scores of 1). Hence, results from our analysis appear to go beyond substitution matrix data.

The empirically conserved substitution groups are consistent with biochemical intuition. The substitution group F Y is the most significant in both databases. Both phenylalanine and tyrosine have side chains with a single aromatic ring, and have similar volume. The group I V contains amino acids with aliphatic side chains that branch at the beta carbon. The groups D N and E Q are both acid-amide combinations with very similar side chains. In addition, the two acidic amino acids, D E, are conserved empirically. However, the two amides, N Q, do not form a substitution group empirically, even though they might seem to belong together on theoretical grounds. As we shall see later, glutamine tends to cluster more with the long-chain polar amino acids, such as lysine and arginine. The basic amino acids, K R, are conserved empirically, and both have amino groups. The group S T contains amino acids that have short hydroxy side chains. Serine is also conserved empirically with alanine (A S); both amino acids are small, each containing a single carbon atom in its side chain. Nevertheless, the smallest amino acid, glycine, does not form a group with serine or alanine, perhaps because glycine has many distinctive properties. Finally, the group H Y is conserved empirically in both databases. Both amino acids have polar ring structures, so the combination of similar volume and polarity appears to account for their conservation.

For amino acid groups of size 3, the two databases identified six empirically conserved substitution groups in common. The highest scoring amino acid triplet in both studies was I L V. All three amino acids in this group have branched aliphatic side chains. As we noted previously, isoleucine and valine form a substitution group themselves. One explanation may be that isoleucine and valine are both branched at their beta carbons, whereas leucine is branched at its gamma carbon. Apparently, branch position matters in some biochemical environments, but not in others. Another conserved amino acid triplet is F W Y. All three amino acids in this group have aromatic side chains, although tryptophan has a double ring. Since phenylalanine and tyrosine themselves form a substitution

| Substitution Group | Pos | C(A) | I(A) | Sep | Pos-induced | Neutral | Neg-induced |
|---|---|---|---|---|---|---|---|
| •FY | 3735 | 183.6 | 74.0 | 109.6 | LWHIVMK | | TSCRQNEPADG |
| •DE | 5980 | 153.0 | 70.0 | 83.0 | KQNSHTAR | | PGMWLYCVFI |
| •KR | 6453 | 157.3 | 93.0 | 74.3 | QEHNSTD | | PAMYLWFVCGI |
| •IV | 10192 | 232.2 | 188.7 | 43.5 | LMTFA | | YCKHQWSNPERDG |
| •ST | 7017 | 105.1 | 62.3 | 42.8 | ANKQED | HVP | MCRYFWILG |
| •AS | 8304 | 91.3 | 70.3 | 21.0 | TKNQEGD | PH | VCMRWFYLI |
| •DN | 4435 | 102.5 | 87.0 | 15.5 | EKSQHTG | R | PAYMWFCVLI |
| •HY | 1728 | 57.3 | 43.3 | 14.0 | FKRQNLW | EV | SDMTCAIPG |
| •EQ | 4856 | 104.3 | 98.9 | 5.4 | KDHRNST | AP | MWYLFCVGI |
| •ILV | 5328 | 188.7 | 86.3 | 102.4 | MFTAYK | H | QCSPWNREDG |
| •FLY | 1474 | 74.0 | 33.9 | 40.1 | IVHMKWT | SR | QCENAPDG |
| •EKQ | 2411 | 85.3 | 49.4 | 35.9 | RDHNTSPA | | MWLYFGCVI |
| •AST | 3293 | 62.4 | 30.7 | 31.7 | KNEQVH | DPM | CFRLWYIG |
| •KQR | 2404 | 83.0 | 65.0 | 18.0 | EHSNTDP | | AYMLVFWGCI |
| FHY | 748 | 40.4 | 28.0 | 12.4 | LKRQWV | NSM | TECIADPG |
| •FWY | 527 | 49.7 | 37.5 | 12.2 | LHM | EKQIRV | DTCNASPG |
| •ILMV | 1696 | 86.3 | 58.8 | 27.5 | FTAYKH | Q | CSWNREPDG |
| FILV | 1502 | 61.9 | 48.1 | 13.8 | MYTAKH | CS | WNQDERPG |
| •EKQR | 1277 | 49.4 | 37.4 | 12.0 | HSDTNPA | | LYFMVWGCI |
| DEKQ | 1173 | 46.1 | 35.6 | 10.4 | NHRSTPA | | WLGYFMVCI |
| HKQR | 746 | 41.4 | 31.2 | 10.2 | ESNTDP | YA | LGVFMWIC |
| FHLY | 372 | 28.0 | 24.3 | 3.7 | KRVQNM | IWETS | CADPG |
| •FILMV | 688 | 48.1 | 26.4 | 21.7 | YATHK | WSCQ | NERDGP |
| •FILVY | 509 | 32.0 | 22.0 | 10.0 | TMHKSW | EC | AQNRDPG |
| EHKQR | 484 | 31.2 | 25.7 | 5.5 | SNDTPA | Y | LFMGVWIC |
| DEHKNQ | 305 | 30.1 | 23.9 | 6.2 | STRPA | GY | FMCVWIL |
| EHKQRS | 351 | 25.7 | 20.5 | 5.2 | NTDPAY | | FGMVLWIC |
| •FILMVY | 260 | 21.2 | 18.1 | 3.1 | THKWS | EANQC | RPGD |
| ADEGHKNPQRST | 63 | 19.4 | 14.6 | 4.8 | Y | VMIL | FWC |
| ADEGHKNPQRSTY | 41 | 6.3 | 2.8 | 3.5 | | LIMVF | WC |

**Table 3** Substitution groups conserved empirically in the BLOCKS database. Groups are arranged according to their size and sorted by separation score. Groups conserved empirically in both BLOCKS and HSSP databases are marked with a bullet. For each group, the table lists the number of positions in the database satisfying the group; the compactness, interference, and separation scores; and their effect on amino acids outside the group, listed in order of descending Z-score.

| Substitution Group | Pos | C(A) | I(A) | Sep | Pos-Induced | Neutral | Neg-Induced |
|---|---|---|---|---|---|---|---|
| •FY | 45082 | 654.9 | 353.7 | 301.2 | WHLMIV | NR | SQTKCAPEDG |
| •IV | 107601 | 933.9 | 731.4 | 202.5 | LMTFA | | YWQRECKHSPNDG |
| •DE | 92101 | 576.0 | 381.6 | 194.4 | NQKSAPTGR | | HMWYCFVIL |
| •ST | 116853 | 474.4 | 290.1 | 106.3 | ANKQEPDR | | HMCVGYWIFL |
| •DN | 84385 | 518.6 | 412.3 | 106.3 | ESQKGTHR | | APYMWCFIVL |
| •HY | 26417 | 269.0 | 172.2 | 96.8 | NFWRQKS | | MDETLAPVCIG |
| •KR | 88140 | 502.6 | 411.6 | 91.0 | QENSTHD | | PAMYGWCVFIL |
| •AS | 121583 | 335.9 | 290.7 | 45.2 | TPENKQDGR | | CMHYVWFIL |
| •EQ | 79815 | 449.6 | 443.4 | 6.1 | KDNRSTAH | | PMYGWCFVIL |
| •ILV | 62336 | 731.4 | 427.2 | 304.2 | MFATY | W | QREKHCSPNDG |
| •FWY | 7941 | 353.7 | 130.3 | 223.4 | LHMRISV | Q | NTCKEADPG |
| •EKQ | 49305 | 433.9 | 267.3 | 166.6 | RNDSTAHP | | MYWGCVFIL |
| •AST | 61564 | 290.1 | 151.6 | 138.5 | NEQKPDR | | VCMGHYWIFL |
| DEN | 46671 | 381.6 | 272.8 | 98.8 | SQKGATRP | H | MYWCFVIL |
| DNS | 50077 | 340.4 | 279.8 | 60.6 | EKQTGARPH | | YMWCFIVL |
| •KQR | 43053 | 372.4 | 315.9 | 56.5 | ENSTHDA | | MPYWGCVFLI |
| NST | 49861 | 260.2 | 216.5 | 43.7 | DKEQRAP | H | GYMCWVFIL |
| •FLY | 19446 | 184.7 | 173.4 | 11.3 | IWMHVR | NS | QTAKCEPDG |
| APS | 30073 | 175.5 | 165.0 | 10.5 | TDEKQNR | | GHCMYWVFIL |
| •ILMV | 19946 | 427.2 | 179.6 | 247.6 | FATYRQ | W | KCHESNPDG |
| FLWY | 3562 | 130.3 | 61.9 | 68.4 | HMIRVQT | S | NKCDAEPG |
| •EKQR | 27456 | 267.3 | 201.4 | 65.9 | NSDTAH | | PMYGWVCFLI |
| DENS | 30264 | 279.8 | 216.9 | 62.9 | KQTAGRP | | HYMWCFVIL |
| FHWY | 2066 | 107.9 | 56.0 | 51.9 | LNRMQS | I | CETKVDAGP |
| DENQ | 25113 | 278.9 | 244.5 | 34.4 | KSRTAHGP | | MYWCFVIL |
| APST | 17550 | 129.1 | 109.7 | 19.4 | EQNKDR | | HGMYVCWFIL |
| AITV | 18933 | 88.6 | 81.6 | 7.0 | LSMEQKR | | YNFHCPWDG |
| •FILMV | 7079 | 179.6 | 124.0 | 55.6 | YTRQWAH | S | KNECPDG |
| •FILVY | 6479 | 180.3 | 134.5 | 45.8 | MWTHARQS | | CNEKPDG |
| FILMY | 3316 | 143.1 | 114.4 | 28.7 | VWRQTH | | ASNCKEPDG |
| ILMTV | 7248 | 77.2 | 56.2 | 21.0 | QRFAKYS | E | WNHCPDG |
| FHLWY | 998 | 56.0 | 35.5 | 20.5 | RMIQNV | D | STECGAKP |
| DEKNQ | 17742 | 237.1 | 222.4 | 14.7 | SRATHP | | GMYWCVFIL |
| AILMV | 7901 | 86.4 | 74.4 | 12.0 | TFRQYE | K | SCPWHNDG |
| DEKNQS | 13239 | 222.1 | 123.8 | 98.3 | RTAPGH | | YMWVCFLI |
| •FILMVY | 2366 | 114.5 | 39.5 | 75.0 | WTQRHAS | | NCKEPDG |
| EKNQRS | 11898 | 167.3 | 157.8 | 9.5 | DTAHG | | PYMWVFCLI |
| FILVWY | 1286 | 68.4 | 59.2 | 9.2 | MQRHT | AS | NECKPDG |
| DEKNQRS | 8083 | 123.8 | 98.2 | 25.6 | TAGHP | | YMWVFLCI |
| DEKNQST | 8238 | 121.2 | 113.4 | 7.8 | ARPGH | | YMVWCLFI |
| ADEKNQST | 5747 | 112.3 | 98.9 | 13.4 | RPGH | | VYMWFICL |
| ADEKNQRS | 5413 | 96.2 | 90.8 | 5.4 | TPHGY | | MVWFLCI |
| FHILMNWY | 102 | 13.2 | 8.5 | 4.7 | VQG | DCPRTA | EKS |
| ADEKNQRST | 3601 | 90.8 | 46.8 | 44.0 | PHGY | MV | FWILC |
| DFGHILMWY | 74 | 22.6 | 9.2 | 13.4 | N | SVTAERPK | CQ |
| ADEKNPQRST | 1283 | 46.8 | 18.8 | 28.0 | HGVL | | MIYFWC |
| DFGHILMNWY | 52 | 9.2 | 2.5 | 6.7 | | TAVPESQCRK | |
| ADEHKNPQRST | 447 | 18.8 | 8.1 | 10.7 | VLY | GMI | FWC |
| ADEHIKLNPQRSTV | 104 | 9.1 | 5.7 | 3.4 | Y | FGM | WC |
| AEFIKLMNPQRSTVY | 42 | 7.8 | 2.0 | 5.8 | | IGDWC | |
| ADEFGHMNPQRSTVWY | 28 | 9.4 | 1.3 | 8.1 | | LCIK | |

**Table 4** Substitution groups conserved empirically in the HSSP database. For explanation, see caption for Table 3.

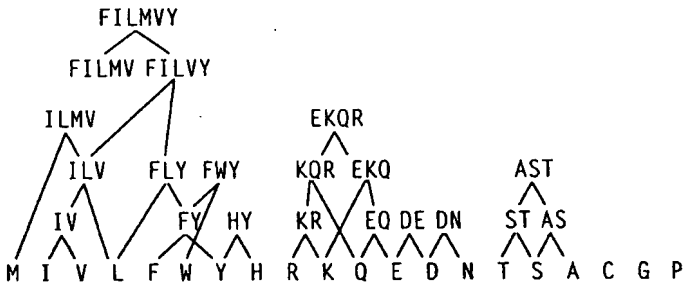



**Figure 3** Classification of empirically conserved amino acid substitution groups. This classification contains all substitution groups that are conserved empirically in the BLOCKS and HSSP databases. The substitution groups are linked by subsumption relationships. Cysteine, glycine, and proline do not belong to any substitution group.

group, it appears that single-ring aromatic side chains are conserved in some contexts, but in other contexts, aromaticity itself is conserved. A closely related amino acid triplet that is also conserved empirically is FLY. The biochemical similarity for this group appears to be volume. Both phenylalanine and tyrosine have a bulky aromatic group at their gamma carbons, whereas leucine has a branched methyl group there. Perhaps in some environments, the branched methyl group provides enough volume and hydrophobicity to substitute for the aromatic ring. The amino acid triplet of AST contains amino acids that have short side chains, with either one or two carbons. The remaining amino acid triplets, EKQ and KQR, contain amino acids with relatively long polar side chains. The biochemical basis for conserving both triplets is not immediately clear. The two groups both contain lysine and glutamine, but one triplet has glutamate and the other has arginine. Arginine can donate a hydrogen bond, whereas glutamate cannot. In addition, the side chain of arginine is much larger than that of glutamate. Perhaps amino acids with long polar side chains and hydrogen bond donor capability (KQR) are conserved in different biochemical environments than amino acids with medium-length polar side chains (EKQ).

The two databases identified only two empirically conserved substitution groups of size 4 in common. One group, ILMV, is a well-recognized group of small hydrophobic amino acids. The other group, EKQR, contains amino acids with long polar side chains. This group subsumes the triplets EKQ and KQR discussed previously. These amino acids have been observed to participate in salt bridges on the surfaces of proteins, which help stabilize protein structure (Goldman 1995).

For substitution groups of size 5, the two databases identified two empirically conserved groups in common. One group, FILMV, contains what are referred to as the major hydrophobic amino acids. The other group, FILVY, demonstrates that tyrosine sometimes acts as a hydrophobic amino acid.

Finally, both databases identified the six-member amino acid group FILMVY as being conserved empirically. This group is a combination of the two substitution groups of size 5 and contains amino acids with hydrophobic characteristics.

For the larger substitution groups, the two databases correlated less well than for smaller substitution groups. Conservation of large substitution groups is difficult to identify because few aligned positions in the BLOCKS and HSSP databases satisfy large conditioning groups. In fact, some of these findings are based on fewer than 100 aligned positions and may not be reliable. Moreover, aligned positions that do satisfy a large conditioning group must contain many sequences and many different amino acids, meaning that the position may not be conserved well. In addition, the large number of sequences means that a few protein families could bias the results. As databases grow larger, we might expect to obtain more accurate results for large groups.

Nevertheless, the two databases showed a near-match with the groups ADEGHKNPQRST in BLOCKS and ADEHKNPQRST in HSSP. The two groups differ on whether glycine belongs, with BLOCKS including glycine and HSSP excluding it. Furthermore, in these larger substitution groups, revealing insights can be obtained from examining the amino acids that they induce negatively. For instance, both databases identify substitution groups that negatively induce the hydrophobic amino acids. These substitution groups differ slightly between the databases, perhaps reflecting the heterogeneity of hydrophilic environments. In addition, both databases identify substitution groups that negatively induce tryptophan, cysteine, and sometimes phenylalanine. These large substitution groups might therefore be defined in a negative sense, by specifying the absence of certain amino acids.

Our set of empirically conserved substitution groups is generally different from those found in the literature, although some similarities can be found. Of the 20 groups found in both BLOCKS and HSSP, eight have been proposed explicitly in previous analyses: DE, KR, ST, DN, ILV, FWY, ILMV, and EKQR. In addition, previous analyses have proposed the groups FILV and ADEGHKNPQRST, which are conserved empirically only in BLOCKS, and the group DENQ, which is conserved empirically only in HSSP.

The twenty substitution groups conserved empirically in both databases can be organized into a classification hierarchy, as shown in Figure 3. In this hierarchy, the amino acids are divided into three major classes. One class, MIVLFWYH, contains hydrophobic amino acids; another class, RKQEDN, contains charged or polar amino acids; and the third class, AST, contains small amino acids. In addition, three amino acids—cysteine, glycine, and proline—do not belong to any substitution group. These amino acids have unique properties that cannot be easily fulfilled by other amino acids. Cysteine can form disulfide bridges. Glycine is the smallest amino acid, having only a hydrogen atom for its side chain. And proline has a distinctive cyclical side chain that causes it to form bends in helices and strands.

## Discussion

Because our approach to amino acid substitution is empirical, it enjoys the same advantages and suffers the same limitations as all empirical studies. One feature of our approach, which could be viewed as either an advantage or limitation, is that our analysis is general. Although our analysis conditions on specific groups that represent specific biochemical contexts, the substitution groups in our study are nevertheless conserved empirically across an entire database. In contrast, many models for describing conservation, such as motifs (Bairoch 1991), profiles (Gribskov, McLachlan, & Eisenberg 1987), and hidden Markov models (Krogh et al. 1994), characterize specific protein families. In those models, each protein family has its own pattern of conservation. The issue is whether generalized biochemical contexts exist and whether selecting aligned positions across an entire database adequately specifies a single biochemical context.

Our opinion is that general patterns of conservation do exist, and that understanding them is critical to understanding specific protein families. We believe that nature is likely to use the same patterns over and over. A strategy of finding general patterns of conservation may be especially fruitful because most protein families are relatively small and the biochemical context of each position is not known. By drawing upon a large amount of data, a general approach is more likely to minimize statistical noise and extract meaningful signals.

Nevertheless, we acknowledge that specific patterns of substitution may occur only in specific protein families. Unfortunately, selecting those protein families may be problematic. An intermediate approach based on secondary structure might prove fruitful, since alpha-helices and beta-strands likely exist in different biochemical environments. Data sets of alpha-helices and beta-strands may generate different sets of substitution groups, which might otherwise be obscured in the entire protein family database. Another strategy would be to analyze large families of proteins with similar function, such as the globins or kinases. In future work, we plan to apply our method to such specialized data sets.

Empirical studies, such as this one, must also consider the issue of sampling bias. Protein families often contain sequences closely related protein sequences that are over-represented, and distantly related sequences that are under-represented. Several methods have been proposed to weight the sequences to remove this bias (Altschul, Carroll, & Lipman 1989; Sibbald & Argos 1990; Vingron & Sibbald 1993). We believe that such weighting methods might help to strengthen our results, and we intend to study the effect of sequence weighting on our analysis of amino acid conservation.

Another characteristic of our work is that our analysis has been biased towards regions that are highly conserved. In particular, the BLOCKS database contains regions with relatively low rates of mutability. Our analysis therefore might miss acceptable substitutions that occur in poorly conserved, highly variable regions. However, our goal has not been to find acceptable substitution groups that are weakly conserved, but rather to identify groups that are strongly conserved and have empirical evidence to support them. We feel that it is these groups that give the best insight into the biochemical principles that are important in protein structure and function.

We anticipate that a set of empirically conserved substitution groups may find several potential applications. First, such substitution groups may provide the basis for new methods for aligning multiple sequences. Most existing methods for aligning multiple sequences rely upon pairwise substitution frequencies. However, substitution groups may provide a more appropriate model for groupwise or consensus relationships. Second, substitution groups might provide an alphabet to describe discrete protein motifs. Most discrete motifs, such as those in PROSITE, are constructed manually, although automated methods have been developed recently (Wu & Brutlag 1995). Both automated and manual methods for building discrete motifs would benefit from having a set of standardized substitution groups. Analogously, our analysis might even provide a basis for probabilistic motifs, such as hidden Markov models (Krogh et al. 1994). The conditional distribution matrix essentially contains probabilistic amino acid profiles for various biochemical contexts. Selected amino acid distributions from the matrix could serve as canonical distributions for hidden Markov models. These distributions might provide general, idealized models of amino acid substitution instead of the empirically tailored distributions based on a specific protein family that are currently used. Recent work on Dirichlet mixture priors (Brown et al. 1993) also tries to find idealized amino acid distributions. Finally, because substitution groups attempt to capture important amino acid properties, they might be helpful in predicting the secondary and tertiary structure of protein sequences. Many researchers have tried to generalize amino acid sequences in terms of their properties (Bork 1989); substitution groups may provide insight into properties that are conserved empirically.

Aside from these applications, though, we hope that our study leads to an improved understanding of amino acid substitution. Amino acid substitution is a central principle in molecular biology. Improved knowledge about amino acid substitution may ultimately lead to better understanding of protein structure and function. Patterns of amino acid substitution represent static evidence of the dynamic process of amino acid evolution and conservation. Findings such as those in this study are central to our understanding of protein structure, function, and evolution.

## Acknowledgments

This work was supported in part by grants LM 05716 and LM 07033 from the National Library of Medicine. The authors thank Tod Klingler for valuable discussions.

## References

Altschul, S. F. 1991. Amino acid substitution matrices from an information theoretic perspective. *J Mol Biol* 219:555–565.

Altschul, S. F., Carroll, R. J., and Lipman, D. J. 1989. Weights for data related by a tree. *J Mol Biol* 207:647–653.

Bairoch, A. 1991. PROSITE: A dictionary of sites and patterns in proteins. *Nucl Acids Res* 19:2241–2245.

Barton, G. J. 1990. Protein multiple sequence alignment and flexible pattern matching. *Meth Enzymol* 183:403–427.

Bork, P. 1989. Recognition of functional regions in primary structures using a set of property patterns. *Febs Letters* 257:191–195.

Brown, M., Hughey, R., Krogh, A., Mian, I. S., Sjölander, K., and Haussler, D. 1993. Using Dirichlet mixture priors to derive hidden Markov models for protein families. In Proceedings of the First International Conference on Intelligent Systems for Molecular Biology, 47–55. Menlo Park, Calif.: AAAI Press.

Dayhoff, M. O., Schwartz, R. M., and Orcutt, B. C. 1978. A model of evolutionary change in proteins. In *Atlas of Protein Sequence and Function*, Nat. Biomed. Research Foundation, pages 345–352.

Goldman, A. 1995. How to make my blood boil. *Structure* 3:1277–1279.

Gonnet, G. H., Cohen, M. A., and Benner, S. A. 1992. Exhaustive matching of the entire protein sequence database. *Science* 256:1443–1445.

Gribskov, M., McLachlan, A. D., and Eisenberg, D. 1987. Profile analysis: Detection of distantly related proteins. *Proc Natl Acad Sci USA* 84:4355–4358.

Henikoff, S. and Henikoff, J. G. 1991. Automated assembly of protein blocks for database searching. *Nucl Acids Res* 19:6565–6572.

Henikoff, S. and Henikoff, J. G. 1992. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 89:10915–10919.

Jimenez-Montano, M. A., and Zamora-Cortina, L. 1981. Evolutionary model for the generation of amino acid sequences and its application to the study of mammal alpha-hemoglobin chains. In Proceedings of the Seventh International Biophysics Congress, Mexico City.

Jones, D. T., Taylor, W. R., and Thornton, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Comp Appl Biosci* 8:275–282.

Kidera, A., Yonishi, Y., Masahito, O., Ooi, T., and Scheraga, H. A. 1985. Statistical analysis of the physical properties of the twenty naturally occurring amino acids. *J Prot Chem* 4:23–55.

Krogh, A., Brown, M., Mian, I. S., Sjölander, K., and Haussler, D. 1994. Hidden Markov models in computational biology: Applications to protein modeling. *J Mol Biol* 235:1501–1531.

Klingler, T. M. 1996. Structural inference from correlations in biological sequences. Ph.D. diss., Program in Medical Informatics, Stanford Univ.

McLachlan, A. D. 1971. Tests for comparing related amino acid sequences. *J Mol Biol* 61:409–424.

Miyata, T., Miyazawa, S., and Yasunaga, T. 1979. Two types of amino acid substitution in protein evolution. *J Mol Evol* 12:219–236.

Mocz, G. 1995. Fuzzy cluster analysis of simple physicochemical properties of amino acids for recognizing secondary structure in proteins. *Protein Sci* 4:1178–1187.

Sander, C. and Schneider, R. 1991. Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins: Structure, Function, and Genetics* 9:56–68.

Sibbald, P. R., and Argos, P. 1990. Weighting aligned protein or nucleic acid sequences to correct for unequal representation. *J Mol Biol* 216:813–818.

Smith, H. O., Annau, T. M., and Chandrasegaran, S. 1990. Finding sequence motifs in groups of functionally related proteins. *Proc Natl Acad Sci USA* 87:826–830.

Smith, R. F. and Smith, T. F. Automatic generation of primary sequence patterns from sets of related protein sequences. *Proc Natl Acad Sci USA* 87:118–122, 1990.

Taylor, W. R. The classification of amino acid conservation. *J Theor Biol* 119:205–218, 1986.

Vingron, M., and Sibbald, P. R. 1993. Weighting in sequence space: a comparison of methods in terms of generalized sequences. *Proc Natl Acad Sci USA* 90:8777–8781.

Vogt, G., Etzold, T., and Argos, P. An assessment of amino acid exchange matrices in aligning protein sequences: The twilight zone revisited. *J Mol Biol* 249:816–831, 1995.

Wu, T. D., and Brutlag, D. L. 1995. Identification of protein motifs using conserved amino acid properties and partitioning techniques. In Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology, 402–410. Menlo Park, Calif.: AAAI Press.

# Molecular evolution of an arsenate detoxification pathway by DNA shuffling

**Andreas Crameri, Glenn Dawes[1], Emilio Rodriguez Jr.[2], Simon Silver[2], and Willem P.C. Stemmer***

*Maxygen, Inc., 3410 Central Expressway, Santa Clara, CA 95051. [1]Affymax Research Institute, 4001 Miranda Ave., Palo Alto, CA 94304. [2]Department of Microbiology and Immunology, University of Illinois, Chicago, IL 60680. *Corresponding author (e-mail: pim_stemmer@qmgates.affymax.com or www.maxygen.com).*

Received 23 October 1996; accepted 3 March 1997

**Functional evolution of an arsenic resistance operon has been accomplished by DNA shuffling, involving multiple rounds of in vitro recombination and mutation of a pool of related sequences, followed by selection for increased resistance in vivo. Homologous recombination is achieved by random fragmentation of the PCR templates and reassembly by primerless PCR. Plasmid-determined arsenate resistance from plasmid pI258 encoded by genes *arsR*, *arsB*, and *arsC* was evolved in *Escherichia coli*. Three rounds of shuffling and selection resulted in cells that grew in up to 0.5 M arsenate, a 40-fold increase in resistance. Whereas the native plasmid remained episomal, the evolved operon reproducibly integrated into the bacterial chromosome. In the absence of shuffling, no increase in resistance was observed after four selection cycles, and the control plasmid remained episomal. The integrated *ars* operon had 13 mutations. Ten mutations were located in *arsB*, encoding the arsenite membrane pump, resulting in a fourfold to sixfold increase in arsenite resistance. While *arsC*, the arsenate reductase gene, contained no mutations, its expression level was increased, and the rate of arsenate reduction was increased 12-fold. These results show that DNA shuffling can improve the function of pathways by complex and unexpected mutational mechanisms that may be activated by point mutation. These mechanisms may be difficult to explain and are likely to be overlooked by rational design.**

Keywords: combinatorial chemistry, sexual PCR, molecular libraries, metabolic engineering

DNA shuffling[1-6] is a process for recursive recombination and mutation, performed by random fragmentation of a pool of related genes, followed by reassembly of the fragments by primerless PCR (Fig. 1). An advantage of DNA shuffling over rational design is that it can optimize the function of genes without first determining which gene product is rate limiting. Our goal is to develop commercial 'molecular breeding' processes for the efficient optimization of complex traits and with minimal understanding of the gene products and their interactions.

This process has been shown to yield rapid functional optimization of genes at rates and to an extent not previously possible in laboratory conditions. However, all previously investigated examples were single genes that were improved by point mutations[2,4-6].

In addition to point mutation we would like to be able to generate diversity using a wide variety of mutational mechanisms, such as polynucleotide deletion, insertion, and inversion, as well as integration and excision. These mutational mechanisms are widely used in nature but find only limited application in rational design.

We have already observed deletions and insertions due to shuffling (data not shown). We show that DNA shuffling is able to access chromosomal integration to obtain rapid functional improvement. The goal of this work was to demonstrate the applicability of DNA shuffling to large, multigene determinants encoding the detoxification of environmentally relevant heavy metals, specifically arsenate.

The trivalent and pentavalent ions of arsenic inhibit many biochemical processes[7]. Vast amounts of arsenate are released in industrial processes, including the mining of arsenopyrite gold ores, which constitute about one-third of world gold reserves. Microbial release of gold from high-arsenic ores is now performed on a commercial scale by microbial oxidation and leaching[8]. Current processes involve incubating large batches of the arsenopyrite ore in tanks for extended periods, during which time a high level of soluble arsenic is released. Untreated, highly toxic arsenic effluent has been disposed of into rivers and has ended up in groundwater[7]. Detoxification of soluble arsenic by bioremediation is a commercially important goal[8]. Our goal was to use molecular breeding to evolve a very efficient pathway for arsenate detoxification, which would have potential for application to bioremediation of arsenate.

## Results and discussion

**Target pathway.** The target DNA used was plasmid pGJ103, a 5.5-kb pUC19-based plasmid containing a 2.7-kb arsenic resistance operon from *Staphylococcus aureus* plasmid pI258 (including the genes *arsR*, *arsB*, and *arsC*), which was expressed in *Escherichia coli*[9]. This operon confers resistance to the toxic oxyanions arsenate (pentavalent arsenic) and arsenite (trivalent arsenic)[7]. This plasmid was evolved by three recursive cycles of DNA shuffling, resulting in point mutagenesis and recombination, and selection for increased resistance to arsenate.

**Resistance evolution.** The wild-type plasmid pGJ103 confers on *E. coli* TG1 resistance to about 4 to 10 mM arsenate when grown on LB plates at 37°C for 24 h. Selection round 1, which was plated on 2 to 32 mM arsenate, yielded about 2,000 colonies growing at 16 mM arsenate. Selection round 2 was plated on 16 to 128 mM arsenate and yielded about 4,000 colonies growing at 64 mM arsenate. Round 3 was plated at 64 to 256 mM arsenate and yielded about 1,500 colonies at 128 mM arsenate. Colonies were harvested from the plates with 128 mM arsenate and replated on 200 to 400 mM arsenate. This gradual increase in resistance during the selection process suggests that more than one mutation is involved.

**Clone AC3.** The largest colony from a plate with 400 mM arsenate from round 3 was found to grow in liquid culture containing up to 500 mM arsenate (Fig. 2B). This clone was called AC3. In addition to arsenate, resistance of AC3 was also increased

by fourfold to sixfold to arsenite ($AsO_2^-$) salts (Fig. 2C), a toxic oxyanion to which resistance requires the ArsB membrane pump but not the arsenate reductase enzyme, ArsC. This suggests mutational effects beyond the *arsC* gene, which only affects resistance to arsenate[9].

**Integration.** Because no plasmid DNA could be recovered from any of the selection round 3 clones, the *ars* operon needed to be recovered by PCR from total cellular DNA. Total cellular DNA of AC3, digested with SacI and SphI, was cloned into pUC19 before sequencing. The operon could only be recovered from AC3 total cellular DNA with primers located at the inner ends of the repeats flanking the operon, which are the terminal portions of site-specific recombinase genes[9]. Primers located in the middle or at the outside of the repeats did not work, suggesting that in AC3 the *ars* operon had integrated into the chromosome via the 200-bp homologous regions. The PCR product containing the evolved operon recovered from chromosomal DNA was reinserted into the native pGJ103 plasmid by mixing with the original pGJ103 plasmid DNA, fragmenting the mixture with DNaseI, reassembling it by primersless PCR, transformating it back into *E. coli*, and selecting for resistance to 400 mM arsenate. Again, the operon appeared to have integrated into the chromosome because in clones resistant to 400 mM arsenate, no plasmid could be detected, and the operon could only be recovered by PCR of chromosomal DNA.

**Selection control without shuffling.** As a control, a similar selection protocol with four cycles of plating on 2 to 32 mM arsenate followed by growth on liquid media was carried out without shuffling. The arsenate resistance remained unchanged at 8 mM, and the plasmid remained episomal. Therefore, some or all of the mutations that resulted from the DNA shuffling were required for the integration and the high-resistance phenotype.

**Mutations.** DNA sequencing of the entire 2.3-kb *ars* operon of AC3 showed 13 base changes relative to the original sequence[9] (Fig. 3). Two silent mutations were found in *arsR* (T389C and T429C), and one was found past the C-terminus of *arsC* (G2469C). Seven silent mutations and three mutations leading to amino acid substitutions were located in *arsB*, the arsenite membrane pump.

The amino acid changes resulted in substitution of hydrophobic amino acids with more hydrophilic residues: mutation T1281C resulted in Leu232Ser, mutation T1317C resulted in Phe244Ser, and mutation T1853C resulted in Tyr423His (numbering as in reference 9). The other seven mutations, T961G, A976G, T1267C, A1402G, T1730C, T1819C, and T1844C, were silent and did not result in predicted amino acid changes.

**Improvement mechanism.** Substitution of hydrophobic with hydrophilic amino acids in a heterologously expressed protein was also found for the evolved green fluorescent protein[5], where it was shown to confer improved protein folding. For the ArsB protein, which is an inner membrane protein whose topology has been experimentally mapped to 12 transmembrane alpha-helical regions in the related plasmid R773 (ref. 10), the explanation may also be improved folding. However, even from this structural model no hypotheses concerning the functional changes can be inferred. For a protein whose overexpression is toxic for cells[10], the mutations may have resulted in a less toxic protein and therefore allowed a higher

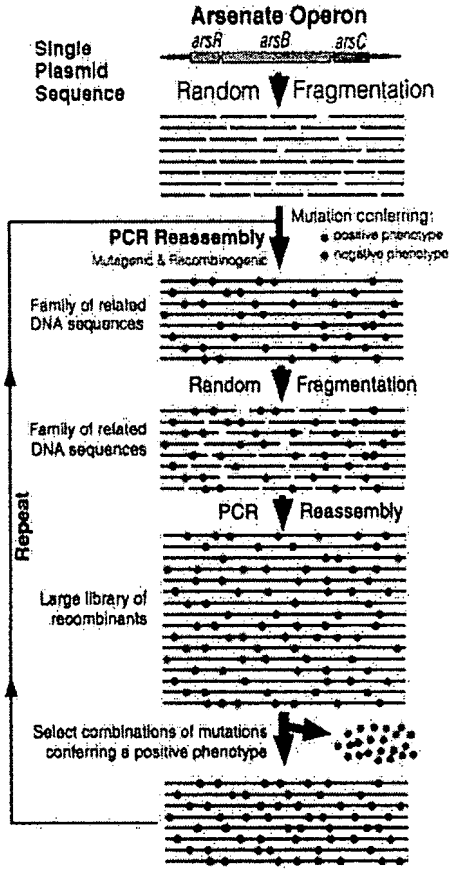



**Figure 1. DNA shuffling. A single sequence is randomly fragmented by DNaseI, and the fragments are reassembled based on sequence similarity by primerless PCR. The diversity consists of the variable level of mutations that is introduced by the PCR reaction. After the first selection cycle, the mutations in the pool of selected sequences are partially unlinked by random fragmentation of the DNA. Following reassembly of the fragments, new combinations of mutations are formed within a conserved framework sequence, and a controlled level of new point mutations is formed as well. When coupled with selection or screening, this process allows rapid accumulation of useful combinations of mutations (green) from multiple parental sequences, while at the same time removing detrimental combinations of mutations (red).**

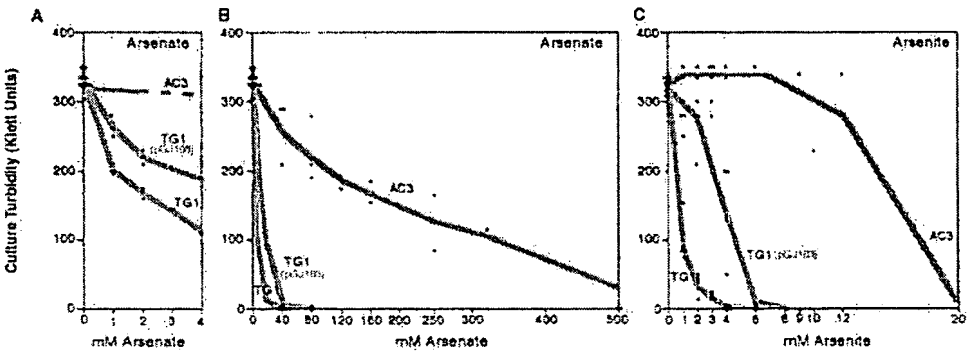



**Figure 2. Resistance to arsenate (A,B) and arsenite (C) before and after DNA shuffling. Arsenate and arsenite resistance for *E. coli* strain TG1 without a plasmid; TG1(pGJ103), containing the wild-type *ars* operon; and AC3, which is TG1(pGJ103) after three rounds of evolution. A composite of four independent experiments.**

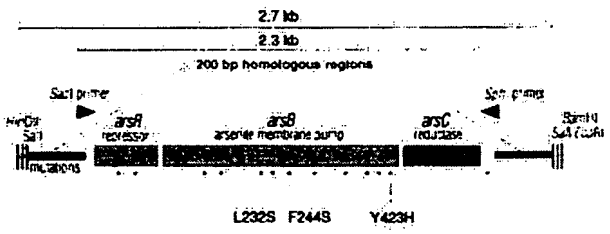



**Figure 3. DNA sequence characterization of the evolved operon. The 13 mutations are marked with asterisks below the boxes representing the three genes.**

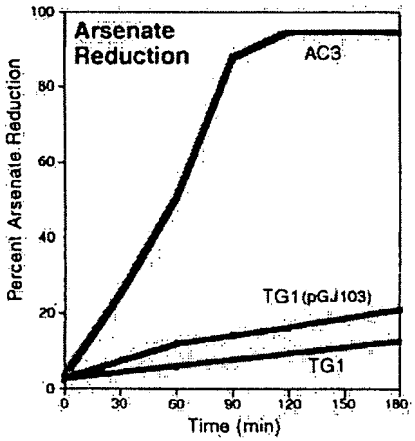



**Figure 4. Whole cell arsenate reduction assay. The ability of *E. coli* constructs to detoxify arsenate was measured by an intact cell arsenate reduction assay[8] using radioactive $^{73}$As-arsenate. Average of two independent experiments.**

expression level resulting in the increased arsenate resistance. An alternative explanation for the increased arsenate resistance is an increased $V_{max}$ of ArsB for arsenate.

**Arsenate reductase activity.** The activity of arsenate reductase by whole evolved AC3 cells was increased about 12-fold (Fig. 4) compared with the control strain with the native pGJ103 plasmid. This increase in whole cell reductase rate appeared to result from an increase in the rate of reduction (V) and not from an enhanced affinity of arsenate for the cells ($K_m$; data not shown). This is consistent with the finding that the mutations occurred in the efflux transport protein (ArsB) and not in the arsenate reductase itself (ArsC). A possible explanation is better coupling between ArsB and ArsC, thereby increasing the arsenate reductase activity even though no mutations are present in the arsC protein.

These results demonstrate that DNA shuffling can dramatically and rapidly improve the performance of a multigene operon, independently of a detailed understanding of the mechanism. For many applications, it will not be necessary to rationalize the improvements mechanistically or a priori. While the shuffling process is simple, like traditional breeding, improvements can arise by a wide variety of complex or simple mechanisms, most of which would have been missed by rational design. We expect that access to a wide variety of mutational mechanisms will generally be important to obtain improvements reliably and rapidly. Some general mechanisms that are likely to be important for sequence evolution in nature, such as the integration strategy we have observed, are not yet a standard part of our current molecular biology toolbox and therefore provide a refreshing view of evolutionary processes.

## Experimental protocol

**DNA shuffling.** The shuffling of pGJ103 was performed as described[3], using 400 to 1,500-bp DNA fragments, which were generated by DNaseI (Sigma, St. Louis, MO) and reassembled by PCR without added primers and using a PCR program as described previously[3]. The PCR product was digested with BamHI and the 5.5-kb fragment was purified from an agarose gel, self-ligated, and electroporated into *E. coli* strain TG1.

**Selections.** Cells were plated on LB plates containing a range of concentrations of sodium arsenate (Sigma) and incubated at 37°C for 24 h. About 1,500 to 4,000 colonies from the plates with the highest arsenate levels were pooled and cultured in liquid LB broth with a similar level of arsenate. Plasmid DNA was prepared from this liquid culture and used for shuffling.

**Selection without shuffling.** As a control, a similar selection protocol was carried out without shuffling. Four cycles of selection on plates with 2, 4, 8, 16, and 32 mM arsenate were performed. DNA (1 μg) was prepared from the most resistant 2,000 to 4,000 colonies and electroporated into TG1 cells without shuffling.

**Cycling.** The second and third shuffling were identical to the first, except that the cells were plated at increasingly higher arsenate levels. After selection round 3 the *ars* operon required recovery by PCR from total cellular DNA due to poor plasmid yield. A 2.3-kb fragment containing the entire *ars* operon but not the 200-bp repeats was recovered from total cellular DNA of AC3 by PCR amplification with oligonucleotide primers (SacI: 5'ACTC-CGTCCACCTGAGCTCTTGGTATT ACACGTCAAACTA [3' A is base 187 in reference 6] and SphI: 5'AACCCTCCCTC GCATGCAAT ATTACAA-GACAG ACAGTT [3' T is opposite base 2512 in reference 9]). The PCR product containing the *ars* operon from clone AC3 was digested with SacI and SphI and cloned into pUC19.

**DNA sequencing.** The entire 2.3 operon from AC3 and from AC5 was sequenced by fluorescent dideoxy termination and with an ABI model 373 (Fig. 3).

**Arsenate resistance measurement.** Cells were grown overnight in LB with 0, 2, 10, or 128 mM arsenate, respectively and were diluted 10,000 times into LB with added oxyanions as indicated. Turbidity was measured after 16 h growth at 37°C. *E. coli* strain TG1 (Pharmacia, Uppsala, Sweden) was used for all experiments. Sodium arsenate was used as a 2.5 M stock solution, neutralized to pH 7 with NaOH.

**Arsenate reduction measurement.** The ability of *E. coli* constructs to detoxify arsenate was measured by an intact cell arsenate reduction assay[11]. Cells were grown as in Figure 1 and were washed and resuspended in triethanol amine buffer[4] and exposed to 4 mM $^{73}$As-arsenate. Samples were removed periodically and heated to 100°C, followed by centrifugation at 4°C, and samples were frozen. Radioactive $^{73}AsO_4^{3-}$ was used as substrate, and $^{73}$As-arsenate and $^{73}$As-arsenite were separated by thin layer chromatography, followed by quantitation in an Ambis radioactive counter.

## Acknowledgments

*We thank Leena Chakravarty for helpful experimental work. This research was supported by Department of Energy grant ER20056.*

1. Stemmer, W.P.C. 1994. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* **91**:10747–10751.
2. Stemmer, W.P.C. 1994. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**:389–391.
3. Stemmer, W.P.C. 1995. Searching sequence space. *Bio/Technology* **13**:549–553.
4. Crameri, A., Cwirla, S., and Stemmer, W.P.C. 1996. Construction and evolution of antibody-phage libraries by DNA shuffling. *Nature Medicine* **2**:100–103.
5. Crameri, A., Whitehorn, E.A., Tate, E., and Stemmer, W.P.C. 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology* **14**:315–319.
6. Zhang, J., Dawes, G., and Stemmer, W.P.C. 1997. Evolution of a fucosidase from a galactosidase by DNA shuffling. *Proc. Natl. Acad. Sci. USA*. In press.
7. Cervantes, C., Ji, G., Ramirez, J.L., and Silver, S. 1994. Resistance to arsenic compounds in microorganisms. *FEMS Microbiol. Rev.* **15**:355–367.
8. Rawlings, D.E. and Silver, S. 1995. Mining with microbes. *Bio/Technology* **13**:773–778.
9. Ji, G. and Silver, S. 1992. Regulation and expression of the arsenic resistance operon from *Staphylococcus aureus* plasmid pI258. *J. Bacteriol.* **174**:3684–3694.
10. Wu, J., Tisa, L.S., and Rosen, B.P. 1992. Membrane topology of the ArsB protein, the membrane subunit of an anion-translocating ATPase. *J. Biol. Chem.* **267**:12570–12576.
11. Ji, G. and Silver, S. 1992. Reduction of arsenate to arsenite by the ArsC protein of the arsenic resistance operon of *Staphylococcus aureus* plasmid pI258. *Proc. Natl. Acad. Sci. USA* **89**:9474–9478.

Plant Physiol. (1996) 110: 835–843

# Nucleotides and Nucleotide Sugars in Developing Maize Endosperms[1]

## Synthesis of ADP-Glucose in *brittle-1*

**Jack C. Shannon*, Fang-Mei Pien, and Kang-Chien Liu**

Department of Horticulture, The Pennsylvania State University, University Park, Pennsylvania 16802–0001

**As part of an in vivo study of carbohydrate metabolism during development of *Zea mays* L. kernels, quantities of nucleotides and nucleotide sugars were measured in endosperm extracts from normal, the single-mutant genotypes *shrunken-1* (*sh1*), *shrunken-2* (*sh2*), and *brittle-1* (*bt1*), and the multiple-mutant genotypes *sh1bt1*, *sh2bt1*, and *sh1sh2bt1*. Results showed that *bt1* kernels accumulated more than 13 times as much adenosine 5′ diphosphoglucose (ADP-Glc) as normal kernels. Activity of starch synthase in *bt1* endosperm was equal to that in endosperm extracts from normal kernels. Thus the ADP-Glc accumulation in *bt1* endosperm cells was not due to a deficiency in starch synthase. ADP-Glc content in extracts of *sh1bt1* endosperms was similar to that in *bt1*, but in extracts of the *sh2bt1* mutant kernels ADP-Glc content was much reduced compared to *bt1* (about 3 times higher than that in normal). Endosperm extracts from *sh1sh2bt1*, kernels that are deficient in both ADP-Glc pyrophosphorylase (AGPase) and sucrose synthase, had quantities of ADP-Glc much lower than in normal kernels. These results clearly indicate that AGPase is the predominant enzyme responsible for the in vivo synthesis of ADP-Glc in *bt1* mutant kernels, but Suc synthase may also contribute to the synthesis of ADP-Glc in kernels deficient in AGPase.**

A number of maize (*Zea mays* L.) endosperm mutants that affect the quantity and quality of carbohydrates in the endosperm, as well as kernel development and morphology, have been identified and extensively studied (Shannon and Garwood, 1984). Mutants such as *waxy* (*wx*), *amylose extender* (*ae*), and *sugary-1* affect the quantity and branching characteristic of kernel polysaccharides (Shannon and Garwood, 1984). Several other mutants involve a defect in the metabolism of sugars, resulting in the accumulation of Suc and the reduction in starch. For example, *sh1* kernels are deficient in the major Suc synthase enzyme (Chourey and Nelson, 1976), and *sh2* and *brittle-2* mutant kernels are both deficient in AGPase (Preiss, 1991). The genetic lesion of *bt1*, another high-sugar/low-starch mutant, is not as well defined. Early screening of developing *bt1* kernels for the activity of various enzymes of carbohydrate metabolism and starch biosynthesis failed to identify a specific enzyme lesion. Developing *bt1* kernels, compared to normal, have been reported to be low in a starch granule-bound phospho-oligosaccharide synthase (Pan and Nelson, 1985) and starch debranching enzyme but twice as high in AGPase (Doehlert and Kuo, 1990).

Sullivan et al. (1991) reported the isolation and analysis of the *Bt1* gene from maize and showed that the protein with greatest similarity to the *Bt1*-encoded protein is a yeast adenylate translocator. Li et al. (1992) showed that the in vitro translated *Bt1* gene product could be imported into chloroplasts, where it was processed and localized to the inner envelope membrane. We have isolated amyloplast membranes from normal and *bt1* kernels and showed that the four most abundant amyloplast membrane polypeptides (38–44 kD) from normal kernels were specifically recognized by antibodies raised against BT1 (Cao et al., 1995). In contrast, the four abundant BT1 immunoreactive polypeptides were missing from amyloplast membranes isolated from *bt1* mutant kernels. Liu et al. (1992) reported that amyloplasts isolated from young maize kernels effectively take up ADP-Glc for starch synthesis, but amyloplasts from *bt1* are only 25% as active in ADP-Glc uptake and incorporation into starch as amyloplasts from normal maize endosperm. The precise function of the *Bt1* gene product is unknown, but it must be important to starch accumulation, since starch content of *bt1* kernels is only about 20% of that in normal maize endosperm (Tobias et al., 1992).

As part of a continuing in vivo/genetic approach to the study of carbohydrate metabolism in normal and starch-deficient maize endosperm mutant genotypes, we discovered that *bt1* kernels accumulate very high levels of ADP-Glc. In this paper we report the nucleotide composition of endosperm extracts from normal and mutant genotypes. Results of a study designed to determine the enzyme(s) responsible for the in vivo synthesis of ADP-Glc in *bt1* kernels are also reported.

[1] This work was supported by U.S. Department of Agriculture Competitive Grant No. 94–37306–0737.

* Corresponding author; e-mail jack_shannon@agcs.cas.psu.edu; fax 1–814–863–6139.

Abbreviations: ADP-Glc, adenosine 5′ diphosphoglucose; AGPase, ADP-Glc pyrophosphorylase; *bt1*, *brittle-1*; DPP, days postpollination; SBE, starch branching enzyme; *sh1* and *sh2*, *shrunken-1* and *shrunken-2*, respectively; UDP-Glc, uridine 5′ diphosphoglucose; UGPase, UDP-Glc pyrophosphorylase; UPTG, UDP-Glc:protein transglucosylase.

## MATERIALS AND METHODS

### Plant Material

The normal maize (*Zea mays* L.) inbred W64A and the endosperm mutant genotypes *sh1* and *bt1* in a near isogenic W64A background were grown in the field at the Russell E. Larson Agricultural Research Farm (Center County, PA) during the summer of 1994. Additional plants for the single mutants *bt1*, *sh1*, and *sh2* and the multiple mutants *sh1bt1* and *sh2bt1*, also in a near-isogenic W64A background, were grown in 20-L plastic pots containing a loose growing medium consisting of two parts peat, two parts perlite, and one part soil. The *sh2* plants were grown in the spring of 1993, and the other single-, double-, and triple- (*sh1sh2bt1*) mutant genotypes were started in the greenhouse in late spring of 1994 and then transferred outside the greenhouse for continued growth. High-intensity sodium lamps were used in the greenhouse to extend the daylength to 16 h. For the experiments comparing nucleotide levels in endosperm tissue containing different dosages of the *Bt1* gene and different developmental ages (10 and 20 DPP), the plants were grown in the greenhouse during the spring of 1993. For the dosage experiment, cross-pollinations and reciprocal cross-pollinations were made between W64A and *bt1*. All ears were hand-pollinated and harvested 20 DPP (except for the developmental age experiment). For plants grown in 1993, kernels were cut from the cobs leaving approximately the lower one-fifth on the cob. The pericarp and embryo were removed from each kernel, and the endosperm was quickly frozen in liquid nitrogen and freeze dried. To reduce the time between ear removal from the plant and freezing, in 1994, the kernels were cut from the ear, quickly frozen in liquid nitrogen, and freeze dried prior to removing the embryos. The dried endosperms and pericarps were pulverized in a mortar and pestle, taking care not to finely pulverize the pericarp. Most of the pericarp was removed by sifting the powdered endosperm through a 40-mesh stainless steel sieve (Cellector Tissue Sieve, E-C Apparatus Corp., Thomas Scientific, Swedesboro, NJ). Although the "endosperm" samples from the kernels harvested in 1994 may have contained some pericarp tissue, in our opinion the advantage of the increased speed of freezing the samples outweighed the disadvantage of the slight pericarp contamination.

### Identification of Double and Triple Mutants by Dot Blot and Embryo Culture

To isolate the double-mutant genotypes, homozygous *sh1* and *sh2* plants were pollinated with pollen from homozygous *bt1* plants. The *sh1*, *sh2*, and *bt1* plants were all in a near-isogenic W64A inbred background. To produce the triple-mutant genotype *sh1sh2bt1*, a *sh1sh2* double-mutant plant (the *sh1sh2* kernels, a gift from William Tracy, University of Wisconsin, Madison, were not isogenic with W64A) was pollinated with pollen from a homozygous *bt1* plant. $F_1$ seeds resulting from these crosses were planted in 20-L containers as described above, and the plants were grown in the greenhouse in the spring of 1994. The $F_1$ plants were self-pollinated and the $F_2$ ears were harvested 20 DPP. Intact kernels were removed from the ears and surface sterilized for 15 min in a 20% chlorine bleach solution containing two drops of Tween 20 per 100 mL of solution. The kernels were rinsed in sterile water, and the upper one-third of at least 100 kernels from each ear was excised for immunoassay by tissue printing. For the immunoassay, the cut surface of each endosperm was blotted onto two (for the double-mutant selections) or three (for the triple-mutant selections) nitrocellulose membranes (see below). The rest of each kernel, including the embryo, was placed in individual wells of a sterile ELISA plate. Each well contained one to two drops of sterile water. The ELISA plates containing the kernel pieces were closed and stored in the cold for about 24 h until the results of the dot blot immunoassay were complete. Care was taken to maintain the identity of the residual kernel pieces with the dot blots. Putative *sh1bt1* kernels were identified by the absence of BT1 antibody-reacting proteins (the BT1 antibody was a gift from Thomas Sullivan, University of Wisconsin, Madison) on one membrane and the absence of Suc synthase (SH1) antibody-reacting proteins (the SH1 antibody was a gift from Prem Chourey, U.S. Department of Agriculture/University of Florida, Gainsville) on a second membrane. Likewise, the *sh2bt1* and *sh1sh2bt1* kernels were identified as those missing the SH2 (antibodies to the SH2 polypeptide were a gift from L. Curtis Hannah, University of Florida, Gainsville) and BT1 polypeptides and those missing the SH1, SH2, and BT1 polypeptides, respectively. Theoretically, one-sixteenth of the $F_2$ kernels on $F_1$ plants from the *sh1* × *bt1* or *sh2* × *bt1* crosses should be double mutants and one-sixty-fourth of the $F_2$ kernels on the *sh1sh2* × *bt1* $F_1$ plants should be the triple-mutant genotype. The embryos from 22 punitive *sh1bt1*, 18 *sh2bt1*, and 3 *sh1sh2bt1* kernels were excised and cultured in vitro according to the procedure of Smith (1992). The cultured embryos were allowed to germinate in the light at 22°C for 14 d. The established seedlings were then removed from the tubes and transplanted to pots containing Fafard Canadian Growing Mix No. 2 (Conrad Fafard Inc., Agawam, MA). After 14 d of growth in the culture room, the seedlings were transplanted to 20-L plastic pots and growth continued in the greenhouse and then outdoors. Each plant was self-pollinated, and two-thirds of the ear were harvested 20 DPP for analysis. The genotype of each ear was confirmed by dot blot immunoassay and by enzymatic analyses, and the remaining one-third of each ear was allowed to develop to maturity. Of all of the plants tentatively identified as multiple mutants, 19 of 22 were *sh1bt1*, 9 of 18 were *sh2bt1*, and 1 of 3 was *sh1sh2bt1*.

The method for tissue printing was adapted from the procedure described by Campillo (1992). Briefly, the nitrocellulose membrane was soaked in 0.2 M $CaCl_2$ for 20 min with shaking and dried on 3-MM Whatman chromatographic paper. Then the membrane was printed with endosperm tissue and blocked in a tray for 1 to 2 h with the blocking buffer (5% [w/v] nonfat dry milk, 1% [w/v] BSA, 1 M Gly, and 0.02% sodium azide in TTBS buffer [20 mM Tris-base, pH 7.5, 150 mM NaCl, and 0.05% Tween 20]) by gentle shaking. The blocking buffer was replaced with

TTBS buffer containing 1% nonfat dry milk, 0.02% sodium azide, and the primary antibody and gently shaken for 1 to 2 h. After the printed membrane was washed twice with TTBS buffer for 5 min each time, it was incubated in the blocking buffer containing the secondary antibody for 1 to 2 h with gentle shaking. The secondary antibody for SH1 consisted of 10 μL of goat anti-mouse IgG-alkaline phosphatase conjugate (Sigma catalog No. A5153) in 10 mL of blocking buffer. For SH2 and BT1, 10 μL of goat anti-rabbit IgG-alkaline phosphatase conjugate (Sigma catalog No. A9919) was added to 10 mL of blocking buffer. The membranes were then washed twice with TTBS buffer for 10 min each time and once with TBS (TTBS buffer without Tween 20) for 10 min. The reaction was then detected with 0.015% (w/v) 5-bromo-4-chloro-3-indoyl phosphate and 0.03% (w/v) nitroblue tetrazolium in AP buffer (0.1 M Tris base, 0.1 M NaCl, and 5 mM $MgCl_2$, pH 9.5). The reaction was stopped by quickly washing the membrane with distilled water. Finally, the membrane was sealed in a plastic bag and stored in a cool place for future reference.

### Enzyme Extraction and Assay

Fifty milligrams of freeze-dried endosperm tissue were homogenized in 2 mL of extraction buffer (10 mM Tricine, pH 7.2, 1 mM DTT, 1 mM EDTA, 0.5 mM PMSF) and then centrifuged at 20,000*g* for 20 min in the cold (4°C). The crude supernatant was used for the assay of AGPase, UGPase, soluble starch synthase, and SBE. The pellet was washed once with extraction buffer and then suspended in extraction buffer, and an aliquot was removed for assay of starch granule-bound starch synthase. Prior to the assay of Suc synthase, an aliquot of the supernatant was desalted by centrifugal filtration on Sephadex G-25 columns equilibrated with extraction buffer. The procedures described by Echeverria et al. (1988) were followed for assay of AGPase, UGPase, starch synthase (including soluble and granular bound), and SBE, except that the activity of SBE was measured using nonradioactive Glc-1-P as a substrate and the released Pi was determined as described by Lanzetta et al. (1979). The Pi released in reactions without added phosphorylase a was subtracted from that released in the complete reaction mixture. Suc synthase activity was measured in the synthetic direction according to the procedure of Guy et al. (1992). Soluble protein in the enzyme extracts was determined using the Bio-Rad protein assay based on the Bradford method. BSA was used as a standard.

### Metabolite Extraction

Equal weights of pulverized, sifted endosperm from three *sh1bt1* ears were combined to produce composite sample 1 and sifted endosperm from another three ears combined for composite sample 2. Likewise, for the two *sh2bt1* composite samples, equal weights of pulverized, sifted endosperm from four and three ears were combined. Pulverized, sifted endosperm from a single ear was used for the *sh1sh2bt1* samples. Pulverized endosperm from several ears were combined for the composite samples of normal (W64A), *bt1*, *sh1*, and *sh2*.

Duplicate 0.5-g samples of freeze-dried endosperm tissues from the single-mutant and triple-mutant genotypes and from the two composite samples of the double-mutant genotypes were homogenized in a 10-mL Potter-Elvehjem-type homogenizer with a Teflon pestle (Thomas Scientific) with 3 mL of 0.8 M $HClO_4$ in a frozen slurry state. The ice slurry was obtained by mixing dry ice powder with the tissue and $HClO_4$ prior to homogenization. Then, the homogenates were immediately centrifuged at 2,500*g* in the cold (4°C) for 10 min, and the pellets were washed twice with cold water (1 mL) each. The combined supernatants (5–6 mL) were collected in 40-mL conical centrifuge tubes and neutralized by addition of an equal volume of cold 0.5 M tri-*n*-octylamine in 1,1,2-trichlorotrifluoroethane. The extract was mixed vigorously for 30 s. Three layers formed during centrifugation at 1,000*g* (4°C) for 5 min. The intermediate layer solidified during centrifugation, and the upper aqueous layer was withdrawn from the tube. Three to 4 mL of $H_2O$ were added to the residual layers, and after the sample was mixed and centrifuged, the upper aqueous layer was removed and added to the initial aqueous extract. This step effectively removes the perchlorate from the aqueous fraction and adjusts the pH to between 6.0 and 6.5 (Heinrich and Rapoport, 1974). The combined aqueous layers were freeze dried for 24 h. The dried extracts were dissolved in water, and any insoluble residues were removed by centrifugation at 16,000*g* for 10 min (Eppendorf centrifuge 5415C) at 4°C. The clarified supernatants were retained for metabolite analysis. In a preliminary study we determined the recovery following extraction of added standards, and all metabolite data were corrected based on the percentage recovery of the individual metabolite.

### Measurement of Neutral Sugars and Starch

Total reducing sugars and Suc in the extracts were determined by the Nelson test (Hodge and Hoereiter, 1962) and the anthrone test (Ashwell, 1957), respectively. To quantify starch granule number and quantity, the pellet remaining after metabolite extraction was suspended in water and aliquots were removed for counting and quantitative measurement. For quantitative measurement, 10-mL aliquots of the starch suspensions were gelatinized for 30 min in a boiling water bath. After adjustment to a known volume, 1-mL aliquots were added to an equal volume of 1.0 M sodium acetate, pH 4.8, containing 10 units of amyloglucosidase. Following an overnight incubation at 37°C, the digests were heated in a boiling water bath for 1 min to inactivate the enzyme and centrifuged at 20,000*g* for 10 min. Glc in the supernatants was determined by the Nelson test (Hodge and Hoereiter, 1962). To determine the number of starch granules, aliquots of the starch suspension were appropriately diluted and the number of granules was determined using a Spencer Bright Line Improved Neubauer hemacytometer (0.1 mm deep) (Hausser Scientific, Blue Bell, PA). The numbers of starch granules in random 0.004-$mm^3$ areas of the hemacytometer were counted using a microscope, and the total number of starch granules per sample extract was calculated.

### Measurement of Nucleotides and Nucleotide Sugars

The major nucleotides (UMP, UDP, UTP, AMP, ADP, and ATP) and nucleotide sugars (UDP-Glc and ADP-Glc) were separated and quantified by ion-paired HPLC using a reversed-phase Adsorbosphere Nucleotide-Nucleoside 7-$\mu$m column (250 × 4.6 mm; Alltech Associates, Deerfield, IL). The mobile phases contained: A, 20 mM $KH_2PO_4$ and 5 mM tetrabutylammonium phosphate, pH 5.0; B, 100% HPLC-grade methanol. The flow rate was 1.5 mL/min. The column was equilibrated with 95% A and 5% B, and then the sample was injected and a linear gradient program from 5% B to 50% B was processed. The 50% B gradient condition was reached in 20 min, and elution continued at the same condition for another 6 min. Each nucleotide was identified by co-migration with commercial standards, and each peak was confirmed by matching the spectrum of each nucleotide peak against that of the standard compounds. The quantity of each nucleotide and nucleotide sugar was determined by comparing the sample peak areas to areas from a standard curve. A Waters Associates HPLC system consisting of dual model 510 pumps, Waters 712 WISP autoinjector, and Waters 996 Photodiode array detector was used for nucleotide analysis. A computer with Waters Millennium 2010 Chromatograph Manager software was connected to the HPLC system to control all of the programs for acquiring and processing the data and calculating the results. The $A_{260}$ was extracted from the photodiode array for quantifying each nucleotide and nucleotide sugar. The relative purity of each peak was determined by comparing the sample nucleotide spectra with standard nucleotide spectra previously stored in the computer. Experiments to determine the effects of *Bt1* gene dosages and developmental ages on nucleotide composition were completed earlier using an older Waters HPLC system, which consisted of Waters dual model 501 pumps, Waters automated gradient controller, Waters model 441 UV absorbance detector (254 nm, 0.2 absorbance units full scale), and the 740 data module.

## RESULTS

### Recovery Study for Extraction Procedure

In the study reported by Tobias et al. (1992), we encountered difficulties associated with excess salt following neutralization of the $HClO_4$ extract with $K_2CO_3$. The residual $KClO_4$ salt in the concentrated extracts limited the size of aliquot that could be used for enzymatic analysis of the hexose-Ps and triose-Ps. Thus, prior to beginning the present study we compared several extraction and neutralization methods. The procedure used in this study was judged to be superior to that used previously (Tobias et al., 1992). The HPLC protocol effectively separated at least 15 compounds that were identified and 5 unidentified peaks (Fig. 1). The identities of the nucleotides were determined by comigration with known standards and confirmed by matching their spectra to those of standards. AMP eluted as the middle peak between two unidentified peaks. The larger unknown peak c has a maximum *A* at 247 nm and its spectrum did not match any of the standard nucleotide spectra. The UDP, ADP-Glc, and ADP peaks in W64A each had leading shoulders. When these nucleotides were quantified, the integrator was set to exclude the area under the shoulders. The elution time and spectrum of minor peak d matched that of GTP, but since it was very low in all genotypes, the quantities of GTP are not reported. Unknown peak e, a contaminant in the eluant, had a maximum *A* at 280 nm. Recoveries for the uridine and adenosine nucleotides, nucleotide sugars, and NAD ranged between 91 and 112% (Table I). Nucleotide quantities reported were corrected by the respective percentage of recovery. Although data concerning the contents of hexose-Ps and triose-Ps are not included in this paper, we note that improved recoveries of these metabolites were obtained using this extraction procedure. For example, with the current procedure we obtained 80 and 71% recoveries of added dihydroxyacetone phosphate and glyceraldehyde-3-P, respectively, a significant improvement over the 66 and 28% recoveries of these metabolites reported by Tobias et al. (1992).

### Enzyme Activities

Preliminary selections of multiple-mutant genotypes were based on immunoassays of endosperm tissue prints from the segregating $F_2$ ears harvested 20 DPP (see "Materials and Methods"). Embryos from putative multiple-mutant genotypes were "germinated" in vitro, and the resulting plants were self-pollinated and ears sampled 20 DPP. Final confirmations of the multiple-mutant genotypes were based on the absence of the endosperm-specific BT1 polypeptide and activities of AGPase and Suc synthase in extracts of the freeze-dried 20-DPP endosperms. Dot blots of extracts from normal, *sh1*, and *sh2* were positive for polypeptides reacting with antibodies raised against BT1, but BT1 antibody-reacting proteins were completely absent from all blots of *bt1*-containing genotypes (data not shown). As expected, Suc synthase and AGPase activities were very low in all *sh1*- and *sh2*-containing genotypes, respectively (Table II). Soluble protein and activities of UGPase, starch synthase (soluble plus granule bound), and total SBE in the mutant genotypes were similar to those in normal kernels.

### Kernel Characteristics of Normal and Single- and Multiple-Mutant Genotypes

Mature kernel phenotypes of the multiple-mutant genotypes are indistinguishable from *bt1* or *sh2* single-mutant kernels. Kernels of the double- and triple-mutant genotypes are viable. In this study we did not measure the endosperm dry weights of the various genotypes. However, we did determine the dry weights per endosperm of these same genotypes grown in the greenhouse in the late winter and spring of 1995 and harvested 20 DPP. The average dry weights per endosperm for normal and *sh1* were similar at 42.3 ± 5.9 and 41.8 ± 2.5 mg, respectively. Weights per endosperm of *sh2*, *bt1*, *sh1bt1*, and *sh2bt1* were similar at 34.6 ± 2.1, 35.0 ± 2.7, 34.7 ± 7.5, and 32.5 ± 3.9 mg, respectively. Although these weights per endosperm are from plants grown in a different year from those ana-

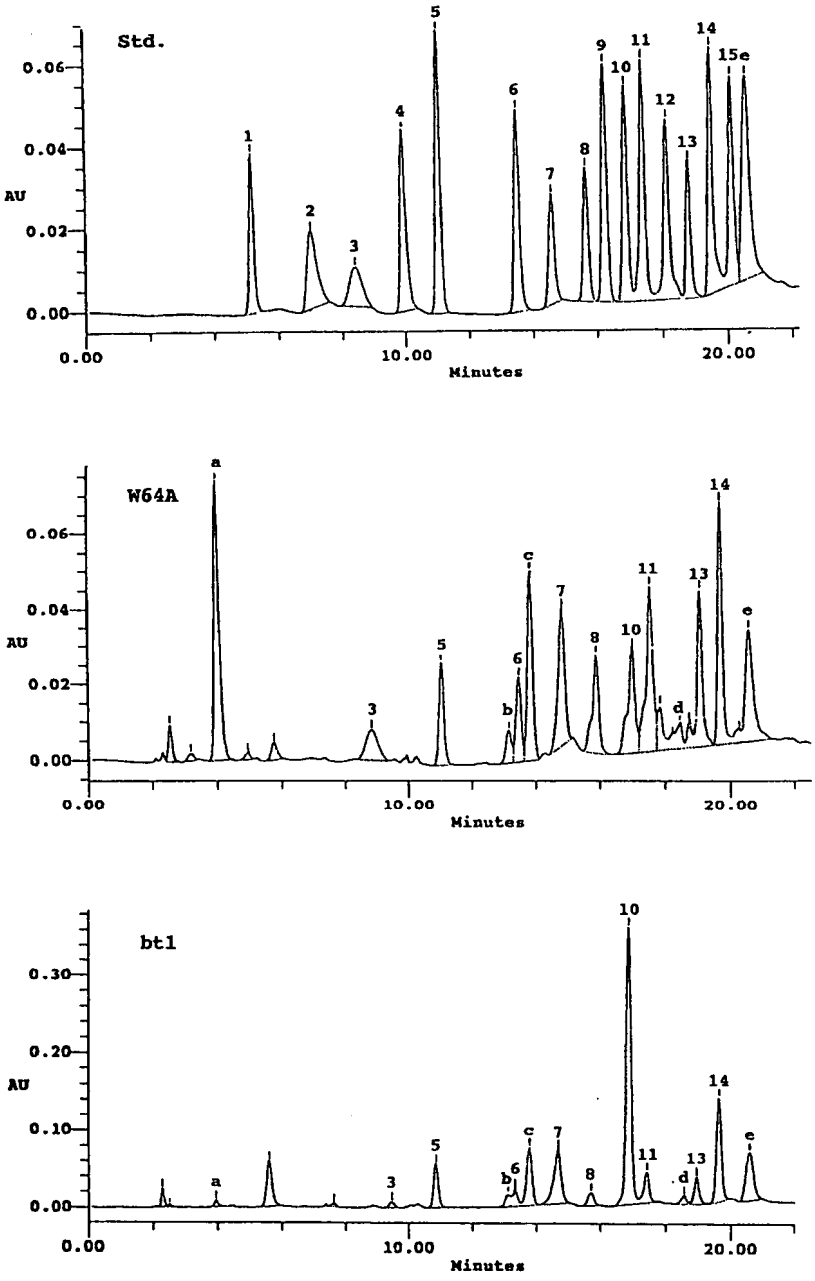



**Figure 1.** HPLC separation of nucleotide standards (top, Std.), endosperm extracts from the normal W64A inbred (middle), and endosperm extracts from the *bt1* mutant (bottom). $A_{260}$ is plotted. Peaks are numbered according to the order of separation of the following standards: 1, uridine; 2, adenine; 3, UMP; 4, adenosine; 5, NAD; 6, AMP; 7, UDP-Glc; 8, UDP; 9, NADP; 10, ADP-Glc; 11, ADP; 12, NADH; 13, UTP; 14, ATP; 15, NADPH. Peaks a to e are unidentified. AU, Absorbance units of $A_{260}$.

lyzed in this study, we are safe in assuming that the relative weights per endosperm of normal (W64A) and the mutant genotypes will be similar, and thus these weights may be used to estimate content per endosperm. Suc and total reducing sugar contents in endosperms of the multiple-mutant genotypes were very high (Table III) as previously reported for the *sh2* and *bt1* genotypes (Tobias et al., 1992). Normal endosperms contained $2.14 \times 10^6$ starch granules $mg^{-1}$ freeze-dried tissue ($90.5 \times 10^6$/endosperm) with an average of 293 $\mu$g starch $10^{-6}$ granules (Table III). The *sh1* endosperm contained about twice as many smaller starch granules as normal ($183.1 \times 10^6$/endosperm), whereas the *sh2*, *bt1*, *sh1bt1*, *sh2bt1*, and *sh1sh2bt1* endosperm samples contained more than 4 times as many very small granules (Table III). The number of starch granules per endosperm for *sh2*, *bt1*, *sh1bt1*, and *sh2bt1* were estimated to be $604.5 \times 10^6$, $466.2 \times 10^6$, $370.2 \times 10^6$, and $312.7 \times 10^6$, respectively.

### Nucleotide Quantity in Normal (W64A) Endosperm Cells

In normal endosperms the sum of uridine nucleotides was 1.7-fold higher than that of adenosine nucleotides. UTP was the highest of all nucleotides, and it was 2.4-, 1.8-, and 1.2-fold higher than UMP, UDP, and UDP-Glc, respectively (Fig. 2). Likewise, ATP was about 2.4-, 1.7-, and

**Table I.** *Metabolite levels in normal (W64A) endosperm (20 DPP) and the percentage recovery of added standards*

One-half gram samples of dry, pulverized endosperm from W64A kernels were extracted, and nucleotides were measured as described in "Materials and Methods." A mixture of nucleotides containing the indicated quantities were added to a second set of 0.5-g endosperm samples prior to extraction. Data are the averages ± SD of two to eight extracts.

| Metabolite | Endosperm Extract | Standard Added | Standard Recovered |
|---|---|---|---|
| | *nmole/g dry wt* | *nmol/g dry wt* | *%* |
| AMP | 452 ± 49 | 400 | 106 ± 10 |
| ADP | 651 ± 93 | 600 | 112 ± 5 |
| ATP | 1097 ± 62 | 1200 | 91 ± 5 |
| ADP-Glc | 596 ± 29 | 600 | 97 ± 3 |
| UMP | 672 ± 57 | 800 | 93 ± 5 |
| UDP | 930 ± 62 | 1000 | 105 ± 2 |
| UTP | 1636 ± 92 | 1600 | 97 ± 5 |
| UDP-Glc | 1416 ± 233 | 1400 | 101 ± 11 |
| NAD | 391 ± 22 | 400 | 98 ± 7 |

1.8-fold higher than AMP, ADP, and ADP-Glc, respectively. The quantity of $NAD^+$ was less than AMP (Fig. 2). The quantities of $NADP^+$, NADH, NADPH, cytosine nucleotides, and guanidine nucleotides were either very low or below the level of detection.

### Nucleotide Quantity in Endosperm Cells of Single Mutants

All mutant data are presented as percentages of nucleotide quantities in normal endosperm. The most striking effect of any of the three mutant genotypes was a 13-fold accumulation of ADP-Glc in the *bt1* mutant endosperm (Fig. 3). ADP-Glc per endosperm in the normal genotype was estimated to be 25.7 ± 1.3 nmol/endosperm, whereas contents per endosperm in *bt1*, *sh2*, and *sh1* were 271.1 ± 31.6, 12.7 ± 2.4, and 34.7 ± 4.0 nmol, respectively. The *bt1* endosperm extract was 2-fold higher in ATP and NAD and 3-fold higher in UDP-Glc but was lower in UMP, UDP, and UTP than extracts from normal endosperms (Fig. 3). Extracts from *sh2* endosperms were 4-fold higher in UDP-Glc but substantially lower than normal in UMP, UTP, and ADP-Glc. ADP, ATP, and NAD in *sh2* were higher than in normal endosperm. The content of nucleotides in *sh1* endosperm extracts was similar to that in normal endosperm with only ADP and ATP being substantially higher (Fig. 3).

### The Effects of *Bt1* Dosage and Developmental Ages on the Quantity of Nucleotides

With each increase in the number of recessive *bt1* alleles, there was an increase in quantity of ADP-Glc and UDP-Glc, but the greatest increase occurred in homozygous recessive kernels (Table IV). All of the nucleotides in endosperms of W64A and *bt1* 10 DPP were much higher than those 20 DPP except for the ADP-Glc in *bt1*. Extracts from 10-DPP *bt1* endosperms contained 5.5 times more ADP-Glc than their normal counterparts, and by 20 DPP the ADP-Glc content had increased to a level more than 17 times higher than that in normal endosperms (Table IV).

### Nucleotide Quantity in Endosperm Cells of Multiple-Mutant Genotypes

To determine whether Suc synthase or AGPase is responsible for the in vivo synthesis of ADP-Glc in *bt1*, we measured the nucleotides in extracts of *sh1bt1*, *sh2bt1*, and *sh1sh2bt1* endosperms. ADP-Glc in extracts of the *sh1bt1* double-mutant endosperms (deficient in Suc synthase) was similar to that in *bt1* (11.5-fold higher than normal), but in extracts of the *sh2bt1* double-mutant endosperms (deficient in AGPase) it was only about 3 times higher than that in normal endosperms (Fig. 4). Extracts of the *sh1sh2bt1* triple-mutant endosperms contained less than half as much ADP-Glc as extracts from normal endosperms. ADP-Glc contents per endosperm for *sh1bt1* and *sh2bt1* were estimated to be 239.6 ± 13.0 and 63.0 ± 0.7 nmol, respectively. If the kernel weight of the triple-mutant genotype is assumed to be the same as the average weight of the double-mutant genotypes, then the ADP-Glc content of *sh1sh2bt1* endosperms would be 9.4 ± 0.1 nmol/endosperm. By comparison, ADP-Glc contents per endosperm in normal and *bt1* endosperms were 25.7 and 271.1 nmol, respectively. UDP-Glc in extracts from all multiple-mutant genotypes was approximately 4-fold higher than in extracts from normal endosperms. The other uridine nucleotides were similar to or lower than in normal endosperms and, except for ADP-Glc in *sh1sh2bt1*, the adenosine nucleotides and NAD were equal to or higher than those in normal endosperm extracts (Fig. 4).

## DISCUSSION

The mature kernel phenotypes of *bt1* and *sh2* are essentially identical, and the starch, neutral sugars, and hexose-P composition of extracts from developing *bt1* and *sh2* are

**Table II.** *Protein and enzyme activities in normal and mutant maize endosperm*

| Genotype | Protein | AGPase | UGPase | Starch Synthase[a] | SBE | Suc Synthase |
|---|---|---|---|---|---|---|
| | *μg mg⁻¹ dry wt* | | | *nmol min⁻¹ mg⁻¹ dry wt* | | |
| Normal | 46.6 | 10.3 ± 0.6 | 55.0 ± 16.4 | 0.7 ± 0.0 | 132.1 ± 23.2 | 22.2 ± 7.9 |
| *sh1* | 48.5 | 12.3 ± 1.7 | 52.9 ± 20.2 | 1.1 ± 0.0 | 177.8 ± 37.6 | 1.6 ± 0.3 |
| *sh2* | 53.6 | 1.8 ± 0.1 | 53.9 ± 10.4 | 1.7 ± 0.0 | 221.9 ± 37.9 | 21.9 ± 4.0 |
| *bt1* | 49.8 | 12.1 ± 0.3 | 35.2 ± 0.2 | 1.6 ± 0.1 | 179.2 ± 45.1 | 16.3 ± 5.8 |
| *sh1bt1* | 53.6 | 15.2 ± 2.9 | 42.4 ± 0.3 | 1.6 ± 0.2 | 194.4 ± 25.7 | 1.3 ± 0.2 |
| *sh2bt1* | 59.9 | 1.8 ± 0.5 | 55.1 ± 3.6 | 1.6 ± 0.1 | 222.3 ± 33.5 | 19.5 ± 0.4 |
| *sh1sh2bt1* | 57.9 | 1.4 ± 0.2 | 44.6 ± 4.7 | 1.8 ± 0.2 | 207.5 ± 39.9 | 0.7 ± 0.1 |

[a] Starch Synthase, Sum of soluble and starch granule-bound activities.

**Table III.** *Neutral sugars and starch granule number and quantity in extracts from 20-DPP normal (W64A) and mutant genotypes*

Sugar data for the double mutants are the means ± SE of duplicate extractions from the two composite samples, and data for the other genotypes are means ± SE of duplicate extractions from composite samples.

| Genotype | Suc | Reducing Sugars | Starch Granule Number | Starch Content |
|---|---|---|---|---|
| | *nmol/mg dry wt* | | *$10^6$/mg dry wt* | *μg/$10^6$ granules* |
| W64A | 129.7 ± 8.2 | 20.0 ± 0.6 | 2.14 | 293 |
| *sh1* | 674.1 ± 9.8 | 49.2 ± 1.8 | 4.38 | 110 |
| *sh2* | 825.1 ± 16.0 | 37.0 ± 0.9 | 17.47 | 19 |
| *bt1* | 1114.4 ± 23.0 | 48.0 ± 0.9 | 13.32 | 19 |
| *sh1bt1* | 1077.0 ± 58.6 | 48.8 ± 3.7 | 10.67 | 21 |
| *sh2bt1* | 981.8 ± 44.1 | 39.5 ± 2.4 | 9.62 | 25 |
| *sh1sh2bt1* | 1198.8 ± 42.2 | 67.2 ± 0.9 | 9.36 | 20 |

very similar (Tobias et al., 1992). In contrast, ADP-Glc content in *sh2* is reduced relative to normal, whereas that in *bt1* is 13- to 17-fold higher than in normal endosperm extracts (Fig. 3; Table IV). Quantities of the other nucleotides in *sh2* and *bt1* were similar. Two enzymes, AGPase and Suc synthase, are capable of synthesizing ADP-Glc in vitro. The *sh2* mutant is deficient in AGPase (Tsai and Nelson, 1966; Dickinson and Preiss, 1969), and *sh1* is deficient in Suc synthase (Chourey and Nelson, 1976). To determine which enzyme is responsible for the in vivo synthesis of ADP-Glc in *bt1* kernels, we produced the multiple-mutant genotypes *sh1bt1*, *sh2bt1*, and *sh1sh2bt1* and determined the content of the various nucleotides. ADP-Glc in endosperm extracts of the *sh1bt1* double mutant was similar (11.5-fold higher than normal) to that in *bt1*, but the endosperm extract from *sh2bt1* was only about 3 times higher in ADP-Glc than normal endosperm extracts (Fig. 4). Extracts of the *sh1sh2bt1* triple-mutant endosperms contained less than half as much ADP-Glc as extracts from normal endosperms. We conclude from these results that AGPase is the predominant enzyme responsible for the in

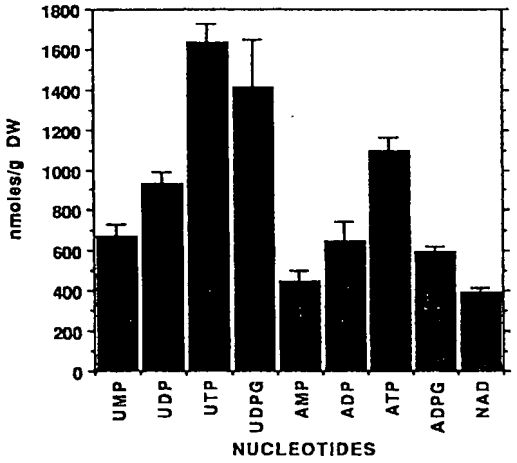



**Figure 2.** Quantity of nucleotides in extracts from normal endosperms (W64A). Data are the means ± SE of four extractions. UDPG, UDP-Glc; ADPG, ADP-Glc; DW, dry weight.

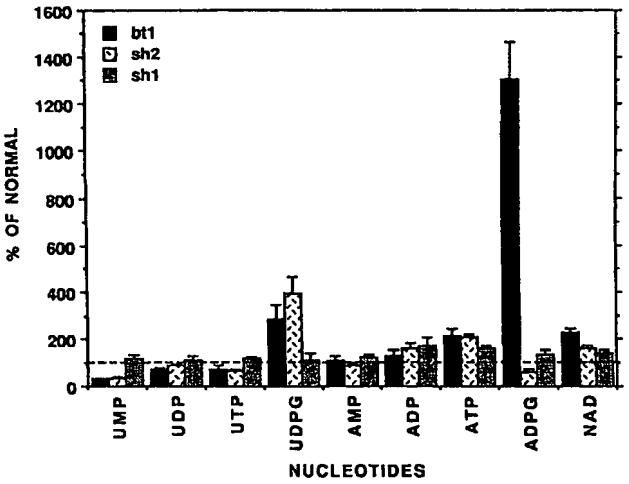



**Figure 3.** Quantity of nucleotides in endosperm extracts from the single-mutant genotypes *bt1*, *sh2*, and *sh1* presented as percentages of the quantity in normal endosperms. UDPG, UDP-Glc; ADPG, ADP-Glc.

vivo synthesis of ADP-Glc in *bt1* mutant kernels. However, it appears that Suc synthase may also play a role in the in vivo synthesis of ADP-Glc, because the extract of *sh1sh2bt1* endosperm, which is deficient in both AGPase and Suc synthase, was much lower in ADP-Glc than extracts of *sh2bt1* or normal endosperms.

AGPase has been reported to be localized in the amyloplast stroma of potato tubers (Kim et al., 1989) and maize endosperms (Miller and Chourey, 1995) and in the stroma of chloroplasts (Preiss, 1991). Localization of AGPase in plastids is generally accepted (Okita, 1992; Smith et al., 1995). Extracts of *bt1* endosperms contain activities of the enzymes for synthesis (AGPase) and utilization (starch synthase) of ADP-Glc that are equal to or higher than those in extracts from normal endosperms (Table II). However, starch accumulation in *bt1* kernels is only 20% of that in normal kernels (Tobias et al., 1992). If ADP-Glc is synthesized in the amyloplast stroma by a plastid-localized AGPase and the activity of starch synthase, another plastid-localized enzyme, is not limiting, then why does the ADP-Glc accumulate rather than being utilized in the synthesis of starch? Synthesis of starch by starch synthase requires the presence of an oligosaccharide primer in addition to ADP-Glc (Preiss, 1991). UPTG (EC 2.4.1.112) is a specialized protein in potato tubers (Lavintman et al., 1974) and maize endosperm (Cura et al., 1994; Rothschild and Tandecarz, 1994; Tandecarz et al., 1995), which functions in the initiation and elongation of the oligosaccharide primer required for de novo synthesis of starch (Cura et al., 1994; Tandecarz, et al., 1995). Maize UPTG is a multimeric protein composed of identical 38-kD subunits. The nuclear-encoded BT1 polypeptides (Sullivan et al., 1991) are of a similar size (39–44 kD) (Cao et al., 1995). BT1 polypeptides have been shown by immunocytochemical localization (Sullivan and Kaneko, 1995) and by studies of isolated amyloplast membranes (Cao et al., 1995) to be specifically localized to the amyloplast membranes of developing maize endosperm. BT1 accounts for about 40% of the

**Table IV.** *The effects of Bt1 gene dosage and kernel age on the quantity of nucleotides in maize endosperm*

Total endosperm (0.5 g dry weight) was extracted and nucleotides measured as described in "Materials and Methods." Values are means ± SE of two replications.

| Endosperm Genotype | Kernel Age | ADP | ATP | ADP-Glc | UMP | UDP | UTP | UDP-Glc | NAD |
|---|---|---|---|---|---|---|---|---|---|
| | *DPP* | | | | *nmol/g dry wt* | | | | |
| *Bt1Bt1Bt1* | 20 | 456 ± 26 | 1525 ± 103 | 482 ± 8 | 428 ± 27 | 924 ± 70 | 1425 ± 139 | 1232 ± 80 | 547 ± 33 |
| *Bt1Bt1bt1* | 20 | 470 ± 10 | 1738 ± 6 | 518 ± 2 | 701 ± 12 | 984 ± 1 | 1400 ± 27 | 1380 ± 6 | 590 ± 2 |
| *Bt1bt1bt1* | 20 | 408 ± 1 | 1695 ± 24 | 581 ± 12 | 568 ± 27 | 906 ± 16 | 1435 ± 41 | 1432 ± 24 | 543 ± 10 |
| *bt1bt1bt1* | 20 | 498 ± 20 | 2273 ± 120 | 8447 ± 294 | 348 ± 4 | 654 ± 24 | 315 ± 20 | 3108 ± 120 | 1212 ± 39 |
| *Bt1Bt1Bt1* | 10 | 1204 ± 50 | 4323 ± 96 | 709 ± 6 | 1058 ± 11 | 1642 ± 26 | 4588 ± 80 | 3514 ± 124 | 690 ± 6 |
| *bt1bt1bt1* | 10 | 1252 ± 98 | 4800 ± 206 | 3932 ± 185 | 1026 ± 49 | 1692 ± 62 | 4027 ± 177 | 4384 ± 180 | 1022 ± 43 |

polypeptides associated with amyloplast membranes from normal endosperms but is missing from amyloplast membranes isolated from *bt1* mutant kernels (Cao et al., 1995). A possible explanation for the ADP-Glc accumulation in *bt1* kernels is that *Bt1* encodes maize UPTG and in its absence de novo starch synthesis is reduced because of a deficiency in oligosaccharide primer. Although this is an attractive possibility, it was not supported by recent evidence from Tandecarz's laboratory. They showed that extracts from endosperms of both W64A (wild type) and *bt1* contained proteins that reacted with antibodies raised against the 38-kD potato tuber UPTG protein (A. Rothschild, S. Bocca, and J. Tandecarz, unpublished results). In addition, our observation of the large number of small starch granules in *bt1* endosperms does not support the suggestion that *Bt1* encodes synthesis of UPTG.

A second possible explanation for the high ADP-Glc in *bt1* endosperm cells is that the ADP-Glc accumulates in a compartment physically separated from starch synthase. Since starch synthase is localized in the plastid stroma (Echeverria et al., 1988), it follows that in *bt1* endosperm cells ADP-Glc accumulates in the cytosol. Kleczkowski et al. (1991) and Hannah et al. (1993) suggested that cereal endosperms may contain AGPase isozymes in the cytosol as well as in the amyloplast stroma, and Villand and Kleczkowski (1994) detailed the evidence in support of this suggestion. A possible explanation is that BT1 is an adenylate translocator, and in its absence (i.e. in *bt1* kernels) ADP-Glc synthesized in the cytosol is unable to enter the amyloplasts and thus it accumulates. This suggestion is supported by the observation that amyloplasts isolated from immature *bt1* endosperm were only 25% as effective in the uptake and incorporation of ADP-Glc into starch as amyloplasts isolated from normal or *wx* endosperms (Liu et al., 1992). In addition, Sullivan et al. (1991) reported that the *Bt1*-encoded protein is most closely related to a yeast adenylate translocator.

A third possible explanation for the ADP-Glc accumulation in the cytosol of *bt1* kernels is that BT1 may be the protein transporter required for the transfer and processing of the AGPase subunits into the amyloplast stroma. According to this suggestion the AGPase subunit transfer would not occur in *bt1* mutant endosperm cells, and instead the AGPase subunit(s) would remain in the cytosol where it is processed into active AGPase. Ballicora et al. (1995) reported that the small subunit of potato AGPase expressed in *Escherichia coli*, without the large subunit, exhibited high enzyme activity, but the regulatory properties differed from those of the heterotetromeric enzyme expressed in *E. coli*. If we assume that the maize amyloplast membrane is unable to transfer ADP-Glc, as in potato tubers (Stark et al., 1992), then accumulation of ADP-Glc in the cytosol of *bt1* endosperm cells would occur. Further studies are needed to conclusively determine the function of BT1.

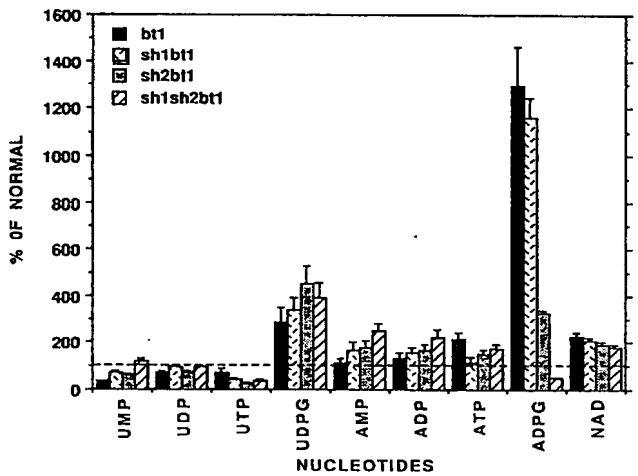



**Figure 4.** Quantity of nucleotides in endosperm extracts from *bt1* and the multiple-mutant genotypes *sh1bt1*, *sh2bt1*, and *sh1sh2bt1* presented as percentages of the quantity in normal endosperms. UDPG, UDP-Glc; ADPG, ADP-Glc.


## ACKNOWLEDGMENTS

We are very grateful to Dr. Thomas D. Sullivan for providing antibodies to the BT1 polypeptide, to Dr. Prem Chourey for antibodies to the SH1 protein, and to Dr. L. Curtis Hannah for antibodies to the SH2 protein.

Received October 16, 1995; accepted December 17, 1995.



## LITERATURE CITED


**Ashwell G** (1957) Colorimetric analysis of sugars. Methods Enzymol 3: 73–105

**Ballicora MA, Laughlin MJ, Fu Y, Okita TW, Barry GF, Preiss J** (1995) Adenosine 5'-diphosphate-glucose pyrophosphorylase from potato tuber. Significance of the N terminus of the small

subunit for catalytic properties and heat stability. Plant Physiol 109: 245–251

**Campillo ED** (1992) Induction, accumulation, localization and transport studies by tissue printing. *In* PD Reid, RF Pont-Lezica, E del Campillo, R Taylor, eds, Tissue Printing, Tools for the Study of Anatomy, Histochemistry, and Gene Expression. Academic Press, San Diego, CA, pp 41–58

**Cao H, Sullivan TD, Boyer CD, Shannon JC** (1995) *Bt1*, a structural gene for the major 39–44 kDa amyloplast membrane polypeptides. Physiol Plant 95: 176–186

**Chourey PS, Nelson OE** (1976) The enzymatic deficiency conditioned by the *shrunken-1* mutations in maize. Biochem Genet 14: 1041–1055

**Cura JA, Tolmasky DS, Lepek VC, Lerner LR, Salerno JC, Krisman CR** (1994) Further studies on the "de nova" process of α1,4-α1,6 glucopolysaccharides—corn starch biogenesis. Cell Mol Biol 40: 1007–1020

**Dickinson DB, Preiss J** (1969) Presence of ADP-Glc pyrophosphorylase in *shrunken* 2 and *brittle* 2 mutants of maize endosperm. Plant Physiol 44: 1058–1062

**Doehlert DC, Kuo TM** (1990) Sugar metabolism in developing kernels of starch-deficient endosperm mutants of maize. Plant Physiol 92: 990–994

**Echeverria E, Boyer CD, Thomas PA, Liu K-C, Shannon JC** (1988) Enzyme activities associated with maize kernel amyloplasts. Plant Physiol 86: 786–792

**Guy CL, Huber JLA, Huber SC** (1992) Sucrose phosphate synthase and sucrose accumulation at low temperature. Plant Physiol 100: 502–508

**Hannah LC, Giroux M, Boyer C** (1993) Biotechnological modification of carbohydrates for sweet corn and maize improvement. Sci Hortic 55: 177–197

**Heinrich R, Rapoport TA** (1974) A linear steady-state treatment of enzymatic chains, general properties, control and effector strength. Eur J Biochem 42: 89–95

**Hodge JE, Hoereiter BT** (1962) Determination of reducing sugars and carbohydrates. *In* RL Whistler, ML Wolfrom, eds, Methods in Carbohydrate Chemistry, Vol I. Academic Press, New York, pp 380–394

**Kim WT, Franceschi VR, Okita TW, Robinson N, Morell M, Preiss J** (1989) Immunocytochemical localization of ADPglucose pyrophosphorylase in developing potato tuber cells. Plant Physiol 91: 217–220

**Kleczkowski LA, Villand P, Lonneborg A, Olsen O-A, Luthi E** (1991) Plant ADP-Glc pyrophosphorylase—recent advances and biotechnological perspectives (a review). Z Naturforsch 46: 605–612

**Lanzetta PA, Alvarez LJ, Reinach PS, Candia OA** (1979) An improved assay for nanomole amounts of inorganic phosphate. Anal Biochem 100: 95–97

**Lavintman N, Tandecarz J, Carcellar M, Mendiara S, Cardini CE** (1974) Role of uridine diphosphate glucose in the biosynthesis of starch. Mechanism of enlargement of a glucoproteic acceptor. Eur J Biochem 50: 145–155

**Li H-M, Sullivan TD, Keegstra J** (1992) Information for targeting to the chloroplastic inner envelope membrane is contained in the mature region of the maize *Bt1*-encoded protein. J Biol Chem 267: 18999–19004

**Liu K-C, Boyer CD, Shannon JC** (1992) Carbohydrate transfer into isolated maize (*Zea mays* L.) amyloplasts (abstract No. 234). Plant Physiol 99: S-39

**Miller ME, Chourey PS** (1995) Intracellular immunolocalization of adenosine 5′-diphosphoglucose pyrophosphorylase in developing endosperm cells of maize (*Zea mays* L.). Planta 197: 522–527

**Okita TW** (1992) Is there an alternative pathway for starch synthesis? Plant Physiol 100: 560–564

**Pan D, Nelson OE** (1985) The deficiency of a starch granule-bound enzyme phosphooligosaccharide synthase in developing *bt1* endosperms. Maize Genet Coop News Lett 59: 105–106

**Preiss J** (1991) Biology and molecular biology of starch synthesis and its regulation. Oxf Surv Plant Mol Cell Biol 7: 59–114

**Rothschild A, Tandecarz JS** (1994) UDP-glucose:protein transglucosylase in developing maize endosperm. Plant Sci 97: 119–127

**Shannon JC, Garwood DL** (1984) Genetics and physiology of starch. *In* RL Whistler, EF Paschall, JN BeMiller, eds, Starch: Chemistry and Industry, Ed 2. Academic Press, New York, pp 25–86

**Smith AM, Denyer K, Martin CR** (1995) What controls the amount and structure of starch in storage organs. Plant Physiol 107: 673–677

**Smith RH** (1992) Plant Tissue Culture, Techniques and Experiments. Academic Press, San Diego, CA, pp 81–86

**Stark DM, Timmerman KP, Barry GF, Preiss J, Kishore GM** (1992) Regulation of the amount of starch in plant tissues by ADP glucose pyrophosphorylase. Science 258: 287–292

**Sullivan TD, Kaneko Y** (1995) The maize *brittle 1* gene encodes amyloplast membrane polypeptides. Planta 196: 477–484

**Sullivan TD, Strelow LI, Illingworth CA, Phillips RL, Nelson OE Jr** (1991) Analysis of maize *Brittle-1* alleles and a defective *Suppressor-mutator*-induced mutable allele. Plant Cell 3: 1337–1348

**Tandecarz JS, Ardila FJ, Bocca SN, Moreno S, Rothschild A** (1995) On the initiation of starch synthesis. *In* HG Pontis, GL Salerno, E Echeverria, eds, Sucrose Metabolism, Biochemistry, Physiology and Molecular Biology: Current Topics in Plant Physiology, Vol 14. American Society of Plant Physiologists, Rockville, MD, pp 107–114

**Tobias RB, Boyer CD, Shannon JC** (1992) Alterations in carbohydrate intermediates in the endosperm of starch-deficient maize (*Zea mays*, L) genotypes. Plant Physiol 99: 146–152

**Tsai C, Nelson OE** (1966) Starch-deficient maize mutant lacking adenosine diphosphate pyrophosphorylase activity. Science 151: 341–343

**Villand P, Kleczkowski LA** (1994) Is there an alternative pathway for starch biosynthesis in cereal seeds? Z Naturforsch 49c: 215–219